DOCUMENT RESUME

ED 130 789 PS 008 931

AUTHOR Seitz, Victoria; And Others
TITLE Long-Term Effects of Intervention: A Longitudinal Investigation.
INSTITUTION Yale Univ., New Haven, Conn.
SPONS AGENCY National Inst. of Mental Health (DHEW), Rockville, Md.; Office of Child Development (DHEW), Washington, D.C.
PUB DATE Sep 76
GRANT NIMH-MH-03008; OCD-90-C-912
NOTE 67p.; Not available in hard copy due to marginal legibility of original document

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS Academic Achievement; Achievement Tests; Compensatory Education; *Early Childhood Education; Educational Policy; *Elementary Education; *Evaluation Methods; Federal Programs; Intelligence Quotient; Intelligence Tests; *Intervention; *Longitudinal Studies; Measurement Instruments; Minority Group Children; Performance Factors; Post Testing; Program Evaluation; Socioeconomic Background
IDENTIFIERS Connecticut; *Project Follow Through; *Project Head Start

ABSTRACT

This longitudinal study examined two groups of children, boys in one sample and girls in another, to assess whether programs such as Head Start and Follow Through produce lasting measureable effects. Low-income children were examined separately from middle-income children and, within the low-income sample, a separate analysis was made for low-income black children (who comprise the largest single group in the final longitudinal sample). Effects of an extensive intervention program were seen upon the children's academic performance in mathematics, general information scores and Peabody Picture Vocabulary Test IQ Scores. Retesting, four to five years following completion of the intervention, showed these children still performing higher than nonintervention children. Reasons for sex differences are discussed; results from interviews with the children and analyses of their school attendance and grades are presented. (Author/BF)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED130789

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

9/76

Long-Term Effects of Intervention: A Longitudinal Investigation

Victoria Seitz
Yale University

Nancy H. Apfel and Carole Efron
Hamden-New Haven Cooperative Education Center

PS 008931

BEST COPY AVAILABLE

Abstract

A key interest in intervention research is whether such large-scale programs as Head Start and Follow Through produce lasting measurable effects. The present study provided a longitudinal examination of this question. For certain groups of children (boys in one year's sample, girls in a second year's sample), effects of an extensive intervention program were seen upon academic performance in mathematics, upon general information scores, and upon Peabody Picture Vocabulary Test IQ scores. Retesting of these children 4-5 years following completion of the intervention showed these children still performing higher than nonintervention children in these areas. Reasons for the sex differences in findings were discussed, and results from interviews with the children and analyses of their school attendance and grades were presented.

One of the most difficult issues facing those who must make social policy decisions concerning special intervention programs is whether programs such as Head Start and Follow Through do or do not produce lasting measurable benefits. As a recent review of the research on early intervention has noted, "The final question must eventually be asked: Can children be brought to a productive level of functioning by (intervention) programs? If the answer is yes, we keep on doing what we are doing. If the answer is no, the social and educational implications are enormous" (Horowitz & Paden, 1973, p. 385).

There has been no shortage of evaluation studies of intervention programs. During the last half of the 1960's, Stearns (1971) estimates, there were, in fact, more than 1,000 studies of the effects of special intervention programs. The vast majority of these studies, however, as Stearns and other reviewers (Bronfenbrenner, 1974; Hawkridge, Chalupsky, & Roberts, 1968; Horowitz & Paden, 1973) have pointed out, have had such serious methodological shortcomings that they have been virtually uninterpretable in regard to evaluating either the short-term or the long-term effectiveness of intervention. Problems of research in this area are not insuperable, however, and a minority of studies have emerged over the past decade which permit several tentative conclusions to be drawn. Perhaps the most valuable of these studies have been longitudinal studies, several of which have recently been critically reviewed by Bronfenbrenner (1974) and by Ryan (1974).

Among the conclusions which both Bronfenbrenner and Ryan draw, the first can be strongly stated: "it is clearly evident from every project that preschool intervention is effective in producing sub-

stantial gains in IQ that are generally maintained so long as the program lasts" (Bronfenbrenner, 1974, p. 14). The long-term stability of such IQ changes, however, as well as the question of whether there are other kinds of effects, are a different matter. The IQ advantage of experimental children over control children, it is reported, usually fails to persist more than a year following the termination of intervention. If this finding were to be taken as indicating psychology's final answer to the question of long-term effects, the answer would clearly be discouraging both to social policy makers and to theorists concerned with the issue of whether the level of cognitive functioning in children can be modified by environmental means. There are reasons to believe, however, that the issue of long-term effects of intervention must still be considered an open one, and the major purpose of the research described in the present report is to address the task of providing data on this matter.

Perhaps the single most important reason that long-term effects are not well understood presently has been the simple absence of empirical data. In contrast to the many studies of young children enrolled in intervention programs, there have been relatively few which examine children after the termination of the program. In the case of the Head Start program, in particular, the present authors know of no published study examining long-term effects into adolescence. An earlier and much publicized study of long-term effects, the Ohio University-Westinghouse Report, concluded that by the end of the second grade, any gains due to the Head Start experience have disappeared and experimental and control children are performing in a manner indistinguishable from each other and undistinguished in comparison with national norms. Despite its widely publicized results,

however, the Westinghouse Report, as a one-time, cross-sectional study with several serious sampling problems (Campbell & Erlebacher, 1970; Smith & Bissell, 1970) cannot be considered an adequate basis for concluding that Head Start has no long-term effects. As Ryan (1974) suggests, longitudinal information is easier to interpret in providing evidence on this matter.

Some encouraging longitudinal data on the effects of Head Start do presently exist to suggest that, contrary to the general conclusion stated above, treated children continue to maintain their advantage over controls on measures of IQ (Abelson, Zigler, & DeBlasi, 1974) and on academic achievement (Abelson, Zigler, & DeBlasi, 1974; Stanford Research Institute, 1971a, 1971b) if they subsequently receive a continued intervention program which is similar to the Head Start program. In the studies cited, the continuation program was provided by a special "Follow-Through" project. Bronfenbrenner (1974) cautiously assesses such findings as indicating that Head Start and Follow Through programs may provide the exception to the general finding that group intervention projects provide few lasting effects. Assuming that this is so, Bronfenbrenner suggests that "One possible consideration lies in the comprehensive character of both the Head Start and Follow Through programs; that is, they are concerned not only with providing an educational program, but also with meeting the needs of the child and his family in the areas of health and social service" (1974, p. 19).

The Abelson et al. study is of particular interest not only because it supports through careful longitudinal work the conclusion of many other studies concerning effects of Head Start after a two-year period, but also because it extends the available information to a comparison made four years after termination of Head Start. Abelson

et al. report evidence that Head Start effects persist for inner-city children who receive no further intervention for at least two years (at the completion of first grade) only to fade out two years later (at the completion of third grade). In comparison, inner-city children who received Head Start and continued intervention did not show a fade out effect: In third grade, the comparison of children who had received continuous intervention with children who had not showed significant advantages for treated children on academic achievement and social-motivational measures.

Given these findings, it would clearly be of interest to determine whether fade out occurs following termination of the extended Follow Through intervention. One purpose of the present study, therefore, was to continue to examine the children from the Abelson et al. study for five years after the completion of the Follow Through progra

In addition to continuing to study the Abelson et al. cohort, we have also been replicating the study with a second cohort. This second cohort of Head Start graduates to enroll in the Follow Through program have been studied since 1968 and have just completed their seventh-grade year in school. At the end of the third grade year, the entire group of 41 children who had completed five years of intervention were compared with a newly-selected cross-sectional control group of 44 randomly selected children from an inner-city school serving the same neighborhood in which many Follow Through children resided. These children were tested extensively during the first two years following termination of intervention, and have just been retested to obtain information at the point of four years following termination of intervention.

If the entire data base for evaluating the effects of intervention consisted only of the two cohorts just described, there would be cause for concern regarding how such factors as possible self-selection of the experimental samples might be assessed to permit clear interpretation of the study's results. As many reviewers have noted (Bronfenbrenner, 1974; Campbell & Frey, 1970; Campbell & Stanley, 1963), such a problem is a common one in evaluations of special programs. In the present case, however, the data to be collected form part of a larger matrix of supporting information which has been collected over the past nine years and which can be employed to provide independent estimates of the degree to which certain kinds of selectivity may be operating. Considerable test data and information on parental education and occupation exist for both the Head Start and non-Head Start children in the first cohort who began but failed to complete the four-year extended intervention program. Some indication of the nature of effects of attrition during the program can therefore be obtained. Similar information is available for the Head Start and non-Head Start children who entered public schools following the Head Start year. Since the completion of the extended intervention program, attrition has been minimal and thus presents few interpretational problems.

Organization of the Present Report

The following section of this report presents a general overview of the research design. There next follows a section on the nature of selection and attrition effects in the present samples. The next section presents data on the academic test performance of children

after the termination of all intervention. The final section presents data on school grades, school attendance, and the interview measure.

I. Overview of the Research

In the Fall of 1967, the Yale research group began a study of kindergarten-aged children who had just completed Head Start (HS) and were enrolling in a special continuing intervention program, Project Follow Through (FT) in Hamden, Connecticut. At that time, comparison non-Follow Through (NFT) children were also selected. These children, some of whom had had Head Start and some of whom had not, were enrolling in inner-city kindergartens in New Haven. Since the FT project was an integrated school with a considerable socioeconomic and racial mixture in its student body, both middle-income and lower-income, and both black and white NFT children were selected from New Haven schools.

NFT children were obtained by taking one intact kindergarten classroom from each of seven different New Haven schools. Three of these schools served low-income neighborhoods from which most of the low-income FT children had been recruited; three of the schools served middle income neighborhoods in which many of the middle income FT children also lived. The seventh comparison school was located in a racially and socioeconomically integrated neighborhood, and the population of this school was similarly an integrated one. One kindergarten classroom was selected from each of these schools by asking the principal to choose a typical classroom in his or her

school.

Parents of the children in all schools were asked to fill out an information sheet indicating the number of other children in the family, whether the child in the present study had had HS or nursery school experience, and information on parental occupation, education, and income. Such information was obtained for all the FT children and for over 90% of the NFT children. The children were individually tested with the Peabody Picture Vocabulary Test (PPVT) at the beginning and again at the end of the kindergarten year. They also received the group-administered "Screening Test of Academic Readiness" (STAR) at these two times. In the following year, the children enrolling in the FT program were tested in the same manner. No new NFT children were tested along with this second cohort of intervention children.

Some data, particularly school-administered Metropolitan Achievement Test scores, were gathered for these children during their first and second grade years in school. Extensive retesting was not conducted until the end of the children's third grade year (at which time the FT intervention program was terminating for the children). At the time of the third grade retesting there had been considerable attrition in both cohorts of FT children and in the NFT samples. For this reason, new samples of third grade age NFT children were recruited at the third grade testing of each of the two FT cohorts. These two new comparison samples were limited to low-income black children, since the large majority of FT children who completed the program were black.

For the first cohort of children, the new cross-sectional control subjects were selected by taking the third grade classmates of the longitudinal control children who could still be located at two of the original four schools from which low-income control children had been obtained. (One of these two schools was the integrated school described previously.) Additional cross-sectional control subjects were obtained from an inner-city school in another city in Connecticut. This latter group consisted of all children who were of third-grade age and who were completing third grade in the newly chosen NFT school. The number of new cross-sectional subjects added was 74; these children were examined only at this one time--no follow-up analyses have been made on these 74 children.

For the second cohort of children, the new cross-sectional control subjects were selected randomly from all children of third grade age who were attending one of the same schools which had provided longitudinal and cross-sectional control children for the first cohort (the school served an almost all-black, all low-income population of children in New Haven). These children have subsequently been studied longitudinally through the seventh grade. Figure 1 summarizes the number of groups studied in the present investigation and the times at which they have been tested.

Insert Figure 1 about here

The test administered since third grade has been the Peabody Individual Achievement Test (PIAT) (Dunn & Markwardt, 1970), an individually-administered test which requires approximately 45

minutes to administer and which provides separate measures of mathematics, reading recognition, reading comprehension, spelling, and general information, as well as an overall academic achievement measure based upon the sum of the individual parts. The test has the advantage of having recently been standardized upon a reasonably representative cross-section of the school population of the United States. The reliabilities of the subtests are satisfactorily high, raging from .63-.94 at the third grade level and .61-.89 at the eighth grade level. The median test-retest reliability coefficient for the PIAT tests over the grades involved in the present investigation is .78. In addition to the PIAT, the children have also received the Peabody Picture Vocabulary Test (PPVT) (Dunn, 1965), which provides an estimate of general intellectual ability based upon vocabulary knowledge. The norming of the PPVT is considerably less adequate than that for the PIAT, and the test is probably most useful only in detecting extreme cases of deviation from the average.

In addition to academic testing, a number of linguistic and motivational measures have been administered to the Cohort 2 FT and NFT children during fourth grade and to both cohorts of FT and NFT children in the fifth-sixth grade testing. The results of the Cohort 2 linguistic testing have been reported elsewhere (Seitz, 1975); the remaining motivational results will be described in a separate report, as the analyses are presently incomplete.

The children from both cohorts have also been interviewed during the fifth-sixth grade and the seventh-eighth grade testing periods. The interview has requested the child to name his favorite

thing about school and his least favorite thing about school (both questions are essentially forced choices to avoid simple socially desirable answers). The child has also been asked whether he likes school or not (a question which does have a social desirability problem, but which nevertheless has yielded some interesting findings). As an indirect way of determining the child's aspirations for himself and his attitudes towards adults, the children have been asked to name the grownups whom they most admire. They have also been asked what they think they will probably be when they grow up and, separately, what they would most like to be if they could be anything they wanted. In addition to the interview, the child's school grades and attendance record have been obtained whenever possible.

II. Selection and Attrition Effects

Several questions are of major interest for the purposes of the present study. First, one may ask how those children who enrolled in the FI program compared on demographic data and test performance with the NFI children from their same neighborhood whose parents chose to send them to their own neighborhood school. This question will be called the selection question. Secondly, one may ask how those children who were still available for study when they were third graders compared with the children who were lost from the longitudinal study. This question will be called the attrition question.

Selection

Kindergarten data were obtained on 121 FI children in Cohort 1

(68 low-income, 53 middle-income), on 117 FT children in Cohort 2 (65 low-income, 52 middle-income), and 136 NFT children in Cohort 1 (72 low-income, 64 middle-income). The total number of kindergarten children studied was thus 374. All of the children who enrolled at FT at any time during the kindergarten year were included in these samples; almost all of the NFT children enrolling in the seven NFT kindergartens (excepting only a few who were absent on scheduled test days) were included in the sample.

Because subsequent findings in the longitudinal study have indicated significant sex differences, and because selection factors might operate differently for the two sexes, boys and girls have been treated as separate samples in the analyses for selection. Low-income children have also been examined separately from middle-income children; and, within the low-income samples, a separate analysis has been made for low-income black children (who comprise the largest single group in the final longitudinal sample of children).

For each sex and each income group separately, individual t-test comparisons were made to compare the children in each kindergarten classroom with the children from every other individual classroom. Such comparisons were made for demographic data (parental education, the number of siblings, pre-school experience) as well as for test performance data (the STAR and the PPVT). Because of the large numbers of *t*-tests calculated, the significance levels of the tests on each dependent variable were corrected by multiplying the probability obtained *t*-value by the number of comparisons which were made.

In order first to determine whether the two cohorts of FT

children were similar, the results of the comparisons of the Cohort 1 FT children with the Cohort 2 FT children will be considered. These comparisons indicated that, for the low-income groups, the two cohorts were generally comparable. During the first year of the FT program, low-income girls were more likely than other groups of low-income FT children to come from single parent homes (almost half were from single parent homes) while in the second year of the FT program, low-income girls came almost entirely from intact families. This difference was a significant one ($X^2(1)=7.26$, $p<.01$). For low-income boys, the end-of-kindergarten PPVT IQ was significantly higher for the first cohort than for the second cohort ($\bar{X}=99.8$ vs. $\bar{X}=91.2$, $t(64)=2.50$, $p<.05$). No other differences were found in comparing the two cohorts of low-income children.

For the middle-income children, the girls in the two cohorts did not differ significantly on any measure except that the girls in Cohort 1 were slightly younger than those in the second cohort ($\bar{X}=60.2$ vs. $\bar{X}=64.6$, $t(48)=3.15$, $p<.01$). The boys, however, showed several indications of possible selection differences. The middle-income boys of Cohort 1 scored significantly higher than did the middle-income boys of Cohort 2 on the end-of-year STAR IQ ($\bar{X}=116.7$ vs. $\bar{X}=106.0$, $t(46)=2.28$, $p<.05$), on the beginning-of-year PPVT IQ ($\bar{X}=110.9$ vs. $\bar{X}=100$, $t(51)=2.33$, $p<.05$), and on the end-of-year PPVT IQ ($\bar{X}=117.0$ vs. $\bar{X}=108.0$, $t(47)=2.23$, $p<.05$). The fathers of the middle-income boys in the first cohort also had significantly more years of education than did the fathers of the boys in the second cohort ($\bar{X}=16.4$ vs. $\bar{X}=14.4$, $t(50)=2.43$, $p<.05$).

In summary, there was little indication of initial differences between the two cohorts of low-income children (and no evidence of any significant difference for the low income black samples). Among the middle-income children, there was evidence that the boys in Cohort 1 were somewhat brighter and came from homes where the father was somewhat better educated than the boys in Cohort 2.

Turning to the results of the comparisons of the FT children with NFT children, Table 1 presents a summary of these comparisons. (Although comparisons were made among all individual classrooms, the data in Table 1 are pooled so that the overall FT-NFT comparison may be seen more clearly.

Insert Table 1 about here

Among the low-income boys, there were three indications of FT-NFT selection effects. (1) In this sample (as in all the other low-income samples), the FT children were much more likely to have had HS experience than were the control children. (2) The end-of-year IQs on the STAR instrument were unusually high at one NFT school ($\bar{X}$=117.3 for the seven low-income boys studied there); this value was significantly higher than the end-of-year STAR IQ for the Cohort 2 FT boys and for the boys at one of the NFT schools ($\bar{X}$=94.7 and $\bar{X}$=88.3, respectively). When the average was taken across all schools, FT children did not differ from NFT children. (3) For the variable of mother's education, boys at one NFT school had mothers whose education was significantly lower than that of the mothers of the boys in the two FT cohorts ($\bar{X}$=9.3 vs. $\bar{X}$=11.6 and $\bar{X}$=11.5, respectively). As

Table 1 indicates, when data from all NFT school children were pooled, the FT and NFT means differed significantly. When the data were re-analyzed for the black children alone, similar results were found for the variables of Hs experience and mother's education, but the variable of end-of-year STAR IQ was no longer significant in any group comparisons.

For low-income girls only the variable of previous Hs experience was significant. The same was true when the data were re-analyzed for black children alone.

For middle-income boys, the Cohort 2 sample scored significantly lower than did the children at one of the middle-income NFT schools on initial PPVT IQ ($\bar{X}$=100 vs. $\bar{X}$=117.21, respectively, $t(40)=3.316$, $p<.05$). On the end-of-year PPVT IQ measure, the Cohort 2 boys again scored lower than did the boys at this same school as well as scoring lower than the boys at another of the NFT schools ($\bar{X}$=108 vs. $\bar{X}$=126.7, $t(37)=3.75$, $p<.01$; and $\bar{X}$=108 vs. $\bar{X}$=125.4, $t(34)=3.16$, $p<.05$, respectively). No other differences were significant. As Table 1 indicates, when the data for all NFT schools and for the two FT cohorts were pooled, the end-of-year PPVT IQ score was significantly lower for the FT than for the NFT boys. Middle-income FT girls did not differ significantly from middle-income NFT girls on any measure.

In summary, there is clear evidence that low-income FT children differed from NFT in having had the preschool experience of attending a Head Start program. There is evidence that the mothers of FT boys have somewhat more education than mothers of NFT boys. The slight evidence for early IQ differences for the low-income boys can perhaps

be discounted since (a) it involved only one of the four low-income MFT schools; (b) it was obtained on one IQ measure, the STAR, but not on the PPVT; and (c) it was not obtained when the data were re-analyzed for black boys only. Overall, there appears to be no strong evidence of selection biases except for whatever factors might lead parents to enroll their children in Head Start.

For middle-income children there was some indication that boys enrolled in the FT project might have been somewhat less bright than middle-income boys in general. Since this result was obtained on one, but not both, IQ measures, and for one cohort but not the other, the conclusion is not a compellingly strong one. Since later results for these middle-income FT children show that they are doing better than average in their school performance, it is worth noting that, if there was any initial selection bias, it did not favor the FT children, and the conclusion that their attendance at the special project for low-income children has not harmed them academically seems a reasonably secure one.

Attrition

To study attrition, the FT children's early test scores were separated into groups according to whether or not these children remained at FT for four years, and t-test comparisons were made for each variable to compare the survivors with the nonsurvivors. Because of the large number of t-tests made, the significance levels were again adjusted on each variable by multiplying the probability level of the obtained value by the number of t-test comparisons made. As before, the data were examined separately for lower and middle-income children, separately for boys and girls, and separately for

the black children within the low-income groups. Unlike the analysis for the selection question, where only kindergarten data were examined all existing data including first and second grade achievement scores were examined in studying attrition.

Low-income samples. Examination of the FT low-income samples as a whole (i.e., including both black and white children) revealed a relatively high attrition rate of 37% for boys, 51% for girls, with most of the loss occurring at the end of the kindergarten year. Much of this loss was due to the exit of white children from the FT program (approximately 70% of the low-income white children did not remain at the FT program for the full four years). The rate of attrition for black low-income children alone was less than for all low-income children (74% of the black low-income boys and 55% of the black low-income girls present in kindergarten completed the full four years of the FT program). The pattern of loss was also gradual rather than abrupt, without the sudden drop in enrollment at the end of kindergarten seen in the white sample. The patterns of loss were comparable in Cohort 1 and Cohort 2.

In the NFT sample, 27 of the 72 original low-income children remained at the end of third grade. Much of this loss was due, however, to the experimenters' decision to drop one of the NFT schools from the study when the children were in first grade because of the introduction of a special tutoring program into the school. With the deletion of this school, the original low-income NFT sample had 48 children, and the percentage of children who were lost was 44%, a figure roughly comparable to the loss in the low-income FT

samples. Tables 2, 3, and 4 present the results of t-test comparisons between the survivors and nonsurvivors of the Cohort 1 FT sample, the Cohort 1 NFT sample, and the Cohort 2 FT sample, respectively.

Insert Tables 2, 3, and 4 about here

As Tables 2 and 4 indicate, boys who remained in the FT program were more likely to have Head Start experience than were boys who left. This finding was not obtained among black boys alone, however, and may simply reflect the larger attrition among white children. There is some evidence that the Cohort 1 boys who left the FT program were less capable than those who remained, while for Cohort 2 the reverse may have been true. For girls in both cohorts, there is some indication that the girls who left the FT program were less academically capable than those who remained. In the NFT sample, as Table 3 shows, what few indications of differential loss exist also suggest the loss of the poorer, rather than the better, students.

In general, however, the major picture conveyed by Tables 2-4 is that despite the many measures upon which survivors and nonsurvivors were compared, very few differences emerged. (The exception to this generalization--Cohort 2 low-income boys--can probably be ignored because only the black children from this cohort have continued to be studied longitudinally; the NFT comparison group selection for Cohort 2 in third grade are also all black). Furthermore, if these comparisons had been limited to data from the kindergarten year alone, where the testing was supervised by the experimenters

and probably more reliable than school-administered achievement test scores, evidence of any selective attrition would have been even less striking.

In general, it seems reasonable to conclude that the FT children who have been studied in the present investigation were relatively typical of inner-city New Haven children except for whatever factors had led their parents to enroll them in a prior Head Start program. Such factors may be important, but in the present study they did not manifest themselves as differences on the many measures investigated with the single exception of maternal education. Examination of attrition suggests that there has been no large or consistent bias which might make the results of the long-term study suspect.

Middle-income samples. Attrition was severe in the middle-income samples, with only 24% of both the boys and the girls who were present in kindergarten completing third grade in the FT program. Approximately half of the original middle-income sample (57% of the boys, 54% of the girls) withdrew at the end of the kindergarten year, and approximately 20% of the original sample (15% of the boys, 18% of the girls) withdrew at the end of the first grade year. Most of those children who remained into the second grade completed the FT program and have continued in the present study into their seventh and eighth grade years.

Although there was substantial attrition, there were few measurable differences between those middle-income children who completed the program and those who did not. As in the low-income groups, children who remained were more likely to have had preschool experienc

(nursery school rather than Head Start for these children) than were children who dropped out of the program. Ninety percent of the children who remained had had nursery school whereas only approximately 50% of the children who did not complete the program had had nursery school. The remaining differences between survivors and non-survivors of the FT program did not exceed the number expected by chance.

III. Academic Test Performance of Children Following Completion of Intervention

Cohort 1

Early Differences Between the FT and NFT Children in the Longitudinal Sample

Before presenting data for the third through eighth grade performance of the Cohort 1 children, a key question of interest is how these particular samples of children compared when they were tested in kindergarten, first and second grade. (This is not identical to the attrition question just discussed, though it is related.) Table 5 presents a summary of the results of these early testings. Only those differences which were statistically significant are reported in Table 5.

Insert Table 5 about here

As Table 5 indicates, FT children were much more likely than NFT children to have had HS experience. Also, parents of the FT children tended to have had somewhat more education than had parents

of the NFT children, with the difference attaining significance for the boys. FT girls were slightly younger than were NFT girls, a difference which can probably be ignored. The PPVT IQ score upon entrance into kindergarten was higher for FT than for NFT boys and remained so through the first grade testing. These higher PPVT IQ scores for the FT boys have been previously reported by Abelson (1974) and seem to reflect the prior HS experience of the FT boys. None of the groups differed significantly on the STAR IQ measures, nor did they differ on measures of academic achievement in grades 1 and 2.

Third-Eighth Grade Performance

Of the 21 FT boys tested in third grade, 17 were located in sixth grade, and two of the lost children were relocated in eighth grade for a sample of 19. All of the 13 NFT boys were retested in sixth grade, and 12 of the 13 were tested in eighth grade. One irregularity in these data is that one NFT boy refused testing in third grade except on the PPVT. The analyses are therefore based upon an N of 12 for most of the third grade FT-NFT comparisons. Among the girls, all 14 FT girls were tested in third, sixth and eighth grades. All 14 NFT girls were tested in third and sixth grade; two children could not be located in eighth grade, resulting in an N of 12.

Although the fluctuations in sample size were small, the number of children in the sample is also small. For this reason, methods of analysis which assess change in performance across time (e.g. repeated measures analyses of variance, multivariate analyses of variance) are not reported at this time. Such methods necessitate choosing between discarding of subjects for whom there is missing data or estimating

the missing data from existing scores. Two years from now, at the completion of this study, we plan to estimate missing data for subjects if they have been located for sufficient testings to permit some confidence in the stability of their performance. At present we prefer the direct approach of comparing FT and NFT groups by t-tests, even though this loses some of the power to be had from repeated measures analyses. We will cautiously reinterpret these results by making tentative estimates of missing scores based on the data presently available.

All analyses have been based upon raw scores in the following manner: each child's raw score performance on a test was compared with the raw score which he should have earned based upon his year and month in school when tested. This procedure avoids potential problems arising from the fact that some children were tested several months earlier than other children. It is important to point out tha these expected raw scores were based upon the grade level which was appropriate for the child as determined by his age and number of years in school. For example, children who had failed to be promoted at some point in their school history were nevertheless judged according to the level appropriate for their age. Table 6 presents a summary of the significance levels of one-tailed t-test FT-NFT comparisons on each test and at each testing period. Where the direction of the FT-NFT difference favors the NFT children, a dashed line indicates this fact.

Insert Table 6 about here

As may be seen in Table 6, and as previously reported by Abelson et al. (1974), at the completion of the FT program, FT boys scored higher on mathematics, on general information, and on the PPVT IQ than did NFT boys. As Table 6 indicates, this pattern has remained similar through the subsequent sixth and eighth grade retestings.

Because of slight fluctuations in the sample, the question reasonably arises as to whether the reported significance levels might be somewhat altered if the entire original sample had been retested at each time. We have made tentative estimates for each child by comparing him with the remaining children in his group at each test period in which the child was tested. The child's performance was converted into a z-score, and the average of these standard score values across all testings was calculated. (This procedure was followed separately for each test.) The child's missing score for any given test was then estimated by giving him the raw score corresponding to his average z-score, calculated in comparison with the children in his group who took the test which he missed. In this manner, the effects of the loss of an exceptionally capable student or an exceptionally poor student can be offset to some degree.

When such an estimation procedure is applied to the third grade data, filling in the missing NFT child's data, the conclusion is generated that the significance levels for mathematics and general information would have been approximately .07 and .05, respectively. For the sixth grade data, applying this estima-

tion procedure to generate the missing four children's scores, the conclusion is generated that the significance levels for mathematics, general information, and PPVT IQ would have been approximately .20, .02, and .02 , respectively, had all the original children been retested. In the eighth grade, the two lost FT boys were highly capable students, while the lost NFT boy was an erratic performer, good in mathematics and spelling, but very poor in general information and reading comprehension. The effects of such loss were such that the apparent loss of the FT superiority in general information is much more likely simply the effect of losing high scoring FT boys and a low scoring NFT boy. The estimated significance levels for mathematics, general information, and PPVT IQ become .13 , .05 , and .05 , respectively when the missing scores are estimated. The remaining measures--reading recognition, reading comprehension, spelling, and the total PIAT score--did not show a significant FT-NFT difference in any analysis of actual or estimated data. The best conclusion from these data appears to be that, except for some loss in mathematics, the pattern of findings for boys reported by Abelson et al. (1974) has remained essentially unchanged five years following the completion of intervention.

For girls, as Table 6 shows, there is little indication that the FT program produced significant gains. The single exception is in mathematics, where there appears to be a sleeper effect--the appearance of a borderline significant finding after previous nonsignificant results. The two NFT girls who were lost from the sample were very poor performers in their earlier testings. The consequence of such a loss is therefore to make the NFT group in eighth grade appear better than

it probably would appear if all children had been tested. The estimated significance level for the FT-NFT mathematics scores is .03, suggesting that this sleeper effect is a genuine one.

The mean values for the academic performance are depicted in Figures 2-7. For greater ease of interpretation, the graphs depict grade equivalence values rather than raw score discrepancies. The mean PPVT I! scores for the FT versus NFT boys were: 100.0 (SD=11.8) vs. 94.1 (SD=12.4) in third grade; 99.2 (SD=14.6) vs. 88.6 (SD=14.7) in sixth grade; and 97.5 (SD=14.2) vs. 89.5 (SD=16.9) in eighth grade. For the FT versus NFT girls the mean scores were: 89.6 (SD=10.4) vs. 86.3 (SD=13.6) in third grade; 84.1 (SD=14.8) vs. 83.3 (SD=14.1) in sixth grade; and 86.4 (SD=17.6) vs. 92.0 (SD=17.8) in eighth grade.

Insert Figures 2-7 about here

Figure 2 shows the FT effect on mathematics. As mentioned earlier, the lost FT children in eighth grade were above average for their group in mathematics when tested previously. There is probably, therefore, no fade out effect in the data. In Figures 3 and 4, the apparent cross over effect with NFT girls exceeding FT girls is probably not a true crossover. Taking all the existing considerations in mind at once, it appears that the results reported by Abelson et al. in 1974 are essentially identical five years later. There has been no fade out, and there is one indication of a sleeper effect for FT girls in mathematics.

Cohort 2

Preliminary Considerations

While we have not yet had time to fully examine the long-term effects of Head Start experience for the children in the present study, preliminary examination of whether this factor was of sufficient importance to require its consideration in analyzing the Cohort findings led to some surprising results. Despite the very small numbers of children (only 3 of 13 FT girls had not had HS and only 3 of 22 NFT girls had had HS), it was evident that the non-HS girls in both groups performed increasingly poorly over time on several measures whereas the HS graduates remained consistently higher. If the FT and NFT girls are pooled and redivided simply on the basis of prior HS experience (with resulting N's of 13 HS graduates and 22 non-HS graduates), t-test comparisons of these two groups (using two-tailed significance levels) revealed that in third grade the HS graduates were superior in general information ($p=.013$) and PPVT IQs ($p=.05$). In fifth grade, the HS graduates were significantly higher on general information ($p=.002$), and also on the total PIAT score ($p=.039$) and the PPVT IQ ($p=.05$). In seventh grade the two groups differed on general information ($p=.048$) and the PPVT IQ ($p=.05$).

The results for the boys showed no HS effects except for a reversed direction finding favoring non-HS graduates in mathematics. Three of 24 FT boys had not had HS experience, while 4 of 22 NFT boys were HS graduates. A comparison of the 25 HS graduates with the 21 non-HS graduates indicated that the HS graduates scored lower in mathematics in the third grade ($p=.04$) and in seventh grade ($p=.00$

No other comparisons were significant.

Because of these differences, it was clear that HS attendance could not be ignored in evaluating the results for Cohort 2. It was for this reason that the decision was made for the Cohort 2 sample to compare children who had had extensive intervention (HS plus FT) with children who had had no intervention at all. (In Cohort 1, similar comparisons of HS graduates with non-HS graduates regardless of FT experience had yielded no significant effects for either boys or girls, a result consistent with the earlier findings of Abelson et al. 1974).

Third-Seventh Grade Performance

Of the 21 FT Cohort 2 boys tested in third grade, 20 were re-tested in fifth grade and again in seventh grade. (One of these boys, however, in the seventh grade testing refused all but the PPVT and the mathematics test; some analyses in seventh grade are therefore based on an N of 20, others on an N of 19). Of the 18 NFT boys, 17 were tested in fifth grade and again in seventh grade. Among the girls, all 10 FT girls were tested each time. Of the 19 NFT girls, 18 were tested in fifth grade and 17 in seventh grade. The girl who was lost in fifth grade was relocated in seventh grade, so that the two girls lost in seventh grade were lost for the first time. All of the control girls, therefore, participated in at least two test periods.

Table 7 presents a summary of the significance levels of one-tailed t-test FT-NFT comparisons on each test and at each testing period. As before, where the direction of the FT-NFT difference

favors the NFT children, a dashed line indicates this fact.

Insert Table 7 about here

As may be seen in Table 7, the FT boys in Cohort 2 showed no evidence of positive effects. There was, in fact, one significant finding which favored the NFT boys on mathematics in grade seven. All other FT-NFT differences were nonsignificant, with the direction generally favoring the NFT boys.

For the girls, the picture was very different. As Table 7 shows, at the end of grade three, FT girls were significantly superior to NFT girls in mathematics, general information, and PPVT IQ score. Two years later, they not only continued to be superior to NFT girls on these same measures, but also they scored significantly higher on spelling and on the total PIAT score. There is thus some evidence for a sleeper effect for these girls. The seventh grade data do not show any continuation of a sleeper trend, but neither do they suggest much loss or change in the original pattern of results.

Again the question of the effects of subject loss arises. If an estimate is made for the scores of the one missing NFT girl in grade five, the p values become .09 for mathematics, .06 for spelling, .005 for general information, .07 for the total PIAT, and .03 for the PPVT IQ. This pattern continues to indicate the sleeper effect for spelling and the total PIAT score and the persistance of the other significant findings. In grade seven, if estimates are made for the scores of the missing two control girls, the significance levels become .075 for mathematics, .04 for general information, .09 for the

total PIAT score, and .065 for the PPVT IQ score.

Figures 8-13 present the grade equivalence scores for the FT and NFT groups on each test at each testing period. The mean PPVT IQ scores for the FT versus NFT boys were: 84.0 (SD=28.9) vs. 91.5 (SD=13.4) in third grade; 88.4 (SD=8.67) vs. 87.0 (SD=9.39) in fifth grade; and 86.8 (SD=11.5) vs. 89.4 (SD=16.2) in seventh grade. For the FT versus NFT girls, the mean scores were: 95.4 (SD=8.5) vs. 87.2 (SD=7.6) in third grade; 95.2 (SD=15.4) vs. 82.9 (SD=13.5) in fifth grade; and 93.4 (SD=12.6) vs. 84.1 (SD=16.2) in seventh grade.

Insert Figures 8-13 about here

Comparison of the Results from the Two Cohorts and General Discussion

When considering the results from both cohorts, one apparently puzzling inconsistency is that it was the FT boys who were superior to their NFT comparison group in Cohort 1, whereas in Cohort 2, it was the FT girls who were superior to the NFT group. In may be helpful in examining this matter to refer to Table 8, which presents a summary comparison of differences between the sexes within each FT cohort and a comparison of the two cohorts. It may also be helpful to refer back to Table 5 which presents FT-NFT comparisons.

Insert Table 8 about here

As Tables 5 and 6 show, there is some indication that the boys in Cohort 1 (both FT and NFT) were brighter than the girls and that the FT Cohort 1 boys came from homes which were more likely to be intact and where the mother was more educated than was true for the FT girls. There may, therefore, have been a sampling problem, and one which affected both the FT and NFT samples in the case of the IQ data. The fact that the differences in IQ were confined to the PPVT instrument, however, and did not manifest themselves on other tests, including the STAR IQ measure, makes the argument for original sampling differences favoring boys somewhat tenuous. In the second cohort, it is noteworthy that differences between the sexes did not begin to emerge until grade one. It does not, therefore, appear in Cohort 2 that the girls were initially any more capable than the boys, but rather that they were more responsive to the effects of intervention. Perhaps the most likely explanation of the sex differences in the findings for the two cohorts is to be found in the fact that the Cohort 1 children were confined to academic survivors, while the Cohort 2 children were not. Such a fact might well reverse the direction of sex effects found in the two cohorts. For example, girls--who are less likely to be behavior problems in the classroom--may be more likely to be promoted even if their academic performance is of borderline quality whereas boys may be more likely to be retained. The effects of such differential promotion practices would be that boys who academically survive in the first few years of school would be likely to be somewhat brighter than the girls in their classrooms.

Whatever the explanation for the different sex effects in the two cohorts, it is interesting that the consistent finding in both Cohorts 1 and 2 was that the FT group which showed significant differences did so in the areas of mathematics, general information, and PPVT IQ scores. There was also consistency in the failure to find any influence of FT on reading performance.

Perhaps the most important findings in the present investigation have been (a) the apparent absence of a "fade out" effect for those groups of intervention children who were significantly ahead of non-intervention children at the end of the program, and (b) the evidence for increasing differences, or a sleeper effect, for girls after the intervention had ceased. This latter effect is a particularly intriguing one since it is precisely the type of effect hoped for in the original implementation of special intervention programs. The implication is that an intervention program can continue to show its effects even after the children have returned to their regular school programs. This finding is consistent with earlier findings of a persistent effect of Head Start through the kindergarten and first grade years, even if children received no further intervention, but a loss of this effect by third grade unless the children had received continued intervention beyond HS (Abelson, 1974; Abelson et al., 1974). In the third grade through seventh grade performance of the Cohort 2 girls, there appears to be a similar persistence and even an increase in the effects of intervention for at least two years following termination of intervention. Encouragingly, in the present case, there does not seem to be a fading out of effects even four

years after termination of intervention.

There are some plausible reasons to explain the emergence of a deeper effect. The first three grades of school, unlike later grades, are concentrated upon the teaching of basic skills; success in the later elementary school grades presupposes the ability of the child to perform these skills and the tutelage concentrates more upon content than method. The effects of intervention might therefore reasonably be expected to be judged more adequately among children in the later elementary school and junior high grades than during the first few years of school. If it is the case that the Cohort 2 results in the present investigation can more readily be generalized than the results from Cohort 1 (because the Cohort 2 children contain both academic successes and failures), the implication is that girls are more affected by educational intervention than are boys. If this is the case (the authors very cautiously suggest that it is), it may be true because the factors which cause low-income boys to perform poorly in school are less easily affected by school-based intervention than are the factors which cause low-income girls to perform poorly. Such factors as peer pressure, early school failure, and the conviction that academic success is not masculine may mitigate against the success of school-based intervention for boys; for girls such intervention may be effective in instilling early self-confidence and school success which can be continued even after the program has ceased.

Other Findings: Middle-Income Children

Since there is no comparison group of children for the middle-income FT children, it is not possible directly to determine the effects of the program upon these children. Their performance can, however, be compared with national norms of the PIAT, and such comparisons show that the middle-income FT children are consistently above national norms in their performance on all measures. For Cohort 1 children the mean grade equivalency scores in the eighth grade testing were 9.7, 9.1, 9.0, 8.8, 9.5, and 9.2, for mathematics, reading recognition, reading comprehension, spelling, general information, and total PIAT score, respectively, and the mean PPVT IQ score was 110.6. For the Cohort 2 children, the mean grade equivalence scores for the seventh grade testing, in the same order, were 10.1, 8.7, 9.0, 7.6, 9.1, and 8.8, and the mean PPVT IQ score was 116.7. From these scores it is clear that no academic harm appears to befall relatively advantaged children who spend their early elementary school years in a program designed especially to benefit low-income children.

Lest the above findings be cited as unconditional evidence in favor of busing programs, however, it must be pointed out that the present FT program differed in significant ways from the school situations often involved in large-scale, court-ordered busing. Attendance at the FT program was voluntarily chosen for their children by all FT families; the school setting was in a quiet, suburban area; the facilities and programs were strongly influenced by parental

choice (e.g. the use of open classrooms); and the opportunities available to the children, such as field trips and a low student-adult ratio (approximately one adult per ten children in the classroom, with both a teacher and a full-time aide) were unusual in comparison with most low-income and many middle-income schools. With all these caveats, the present data should provide reassuring evidence to middle-income parents that academic retardation is by no means an automatic outcome of their child's attending schools which are heterogeneous in student background. As previously reported, middle-income children attending such a school do not even acquire facility with nonstandard English speech patterns, even though the low-income children become more proficient in standard English (Seitz, 1975). Whether both middle- and lower-income children acquire through this integrated schooling a greater understanding of and social ease with persons from different cultural backgrounds remains an as-yet not well explored question.

IV. Interview Responses, School Grades, and School Attendance

The data to be described in this section were primarily in the form of frequency counts. Chi square analyses, or Fisher's Exact Test in cases with very small Ns, were employed to compare FT and NFT children of each sex within each cohort.

Interview Data

Children's responses to the question of whether they liked school were coded into "yes" versus other answers (e.g. "no", "sometimes"). Among the Cohort 2 girls, the FT children were significantly more likely to report liking school than were the NFT girls (50% versus 89%, $p=.04$ by Fisher's Exact Test). The comparisons for the remaining groups were not significant.

Responses to the question concerning the child's favorite thing about school were coded into the following categories: (a) an academic subject; (b) a nonacademic subject, such as woodworking, shop; (c) social activities; (d) nonspecific academic aspects (e.g. "learning"), and (e) "nothing." Because of small Ns in some categories, these categories were collapsed into categories (a) and (d) (academic aspects) versus (b) and (c) nonacademic aspects). Comparisons of the FT and NFT groups in each sex by cohort group revealed no significant differences. Very few children responded to this question by saying that there was nothing which they liked about school. Interestingly, however, all such children were FT graduates. Pooling across cohorts and sexes, 8% of FT graduates replied that there was no favorite thing about school for them, while 0% of NFT children gave this response ($p=.035$ by Fisher's Exact Test).

Responses to the question which asked the child to name the worst thing about school were coded into the following categories: (a) an academic subject; (b) a nonacademic subject; (c) social aspects (such as children fighting in halls or bathrooms); (d) rules or discipline aspects (such as being sent to the principal's office,

being required to take showers in gym); and (e) nothing. Because of small N's in some categories, these categories were collapsed into (a) academic aspects) versus (c) and (d) (rules, discipline and social problems. Category (b) was very rarely named. Comparisons of FT and NFT groups revealed no significant differences in frequency of naming these two categories. Relatively few children gave a response indicating that nothing was particularly bad about school. Most such children who gave this response were NFT children. Collapsing across cohorts and sexes, only 5% of FT children indicated that there was nothing which they felt to be particularly bad about school, while 18% of the NFT children gave this response ($p=.025$ by Fisher's Exact Test).

The answers to the question of what grownups the child most admired were coded into the following categories: (a) family members, other relatives, or family friends; (b) community leaders; (c) persons in sports or performing arts; (d) teachers; and (e) others. Categories (b) and (e) rarely were mentioned. Analyses were therefore made comparing FT and NFT children with respect to ;whether or not they mentioned family members and friends among adults they most admired, whether or not they mentioned sports and performing arts persons, and whether or not they mentioned teachers. None of these analyses yielded significant results.

The child's career aspiration was graded into socioeconomic status categories on a scale of 1-5 with 1 denoting unskilled, 2 semiskilled, 3 clerical, 4 white collar and managerial, and 5 professional occupations. A similar coding was made of the child's re-

sponse to the question of what he would like to be if he could be anything he wanted. Analyses of variance on both the real and the ideal aspiration levels produced no significant results.

Grades and Attendance

On the basis of the child's school grades and his grade level, each child's academic success was coded as follows: (a) grossly failing, as indicated by the child's having been retained in a lower grade than appropriate for his age or in a special learning disabilities group; (b) at grade level, but failing (earning D's and F's); (c) at grade level with marginal to average performance (earning C's and D's); (d) at grade level with average to good performance (earning C's and B's); and (e) superior (above grade level or at grade level; earning A's and B's). Because of small Ns in some groups, these categories were collapsed into (a) and (b) (failing) versus (c), (d), and (e) (adequate or better school performance). Comparisons of FT and NFT groups in each sex by cohort grouping yielded no significant differences in the frequency of representation of FT and NFT children in these categories.

School attendance was categorized as adequate (defined as 0-15 absences during the school year) versus poor (16 or more absences). Comparisons of FT and NFT children yielded no significant differences in school attendance.

Discussion of Findings from Interviews, Grades, and School Attendance Data

In summarizing the results from the interview, grades, and attendance data, it appears that one effect of the FT program has been to lead children to be more critical of their post-FT school experience. Surprisingly, however, a lesser liking for school does not seem to spell itself out in poorer attendance or school grades. In fact, as previously noted, the FT girls in Cohort 2 have consistently been better academic performers on several measures than have the NFT girls, yet these FT girls are particularly likely to report that they do not like school.

It is difficult to gauge what the long-range effects of such feelings might be. Possibly, if fade out does eventually occur, it may be for this reason. It is also clear from the interview, however, that the intervention children believe that they must be special in some way, and they frequently reminisce positively about their early years at FT. The children's reflections seem, in fact, almost nostalgic. Their present tendency to criticize school appears therefore to reflect an awareness of what a more positive school experience could be. The eventual effects of their attitudes toward school could, of course, be either positive or negative. Their faith in the potentialities of school programs might sustain them through otherwise difficult times. On the other hand, the frustration engendered by the comparison of what might be with what is may eventually lead to a drop in performance. Two years from now, when these same children

are in ninth and tenth grades, we plan to test them for the final time to determine the effects of the early intervention. At that time we will issue a final report of this study.

References

Abelson, W. D. Head Start graduates in school: Studies in New Haven, Connecticut. In S. Ryan (Ed.), A report on longitudinal evaluations of preschool programs. Vol. 1. Longitudinal evaluations. Washington, D. C.: Department of Health, Education, and Welfare, Publication No. (OHD) 74-24, 1974.

Abelson, W. D., Zigler, E., & DeBlasi, C. Effects of a four-year Follow-Through program on economically disadvantaged children. Journal of Educational Psychology, 1974, 66, 756-771.

Bronfenbrenner, U. A report on longitudinal evaluations of preschool programs. Vol. 2. Is early intervention effective? Washington, D. C.: U. S. Department of Health, Education, and Welfare, Publication No. (OHD) 74-25, 1974.

Campbell, D. T. & Erlebacher, A. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), Compensatory education: A national debate. Vol. III of the disadvantaged child. New York: Brunner/Mazel, 1970.

Campbell, D. T. & Frey, P. W. The implications of learning theory for the fade-out of gains from compensatory education. In J. Hellmuth (Ed.), Compensatory education: A national debate. Vol. III of the disadvantaged child. New York: Brunner/Mazel, 1970.

Campbell, D. T. & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1963.

Dunn, L. M. Peabody Picture Vocabulary Test. Minneapolis, Minn.: American Guidance Service, 1965.

Dunn, L. M. & Markwardt, F. C. Peabody Individual Achievement Tests. Circle Pines, Minnesota: American Guidance Service, 1970.

Hawkridge, D. G., Chalupsky, A. B., & Roberts, A. O. H. A study of selected exemplary programs for the education of disadvantaged children. Palo Alto, Calif.: American Institute for Research, 1968.

Horowitz, F. D. & Paden, L. Y. The effectiveness of environmental intervention programs. In B. M. Caldwell & H. N. Riccuiti (Eds.) Review of Child Development Research. (Vol. 3). Chicago: University of Chicago Press, 1973.

Ryan, S. (Ed.), A report on longitudinal evaluations of preschool programs. Vol. 1, Longitudinal evaluations. Washington, D. C.: U. S. Department of Health, Education, and Welfare, 1974.

Seitz, V. Integrated versus segregated school attendance and immediat recall for standard and nonstandard English. Developmental Psychology, 1975, 11, 217-223.

Seitz, V., Abelson, W. D., Levine, E., & Zigler, E. Effects of place of testing on the Peabody Picture Vocabulary Test scores of disadvantaged Head Start and non-Head Start children. Child Development, 1975, 46, 481-486.

Smith, M. S. & Bissell, J. S. Report analysis: The impact of Head Start. Harvard Educational Review, 1970, 40, 51-104.

Stanford Research Institute. Implementation of planned variation in Head Start: Preliminary evaluation of planned variation in Head

Start according to Follow-Through approaches (1969-70). Washington, D. C.: Office of Child Development, U. S. Department of Health, Education, and Welfare, 1971. (a)

Stanford Research Institute. Longitudinal evaluation of selected features of the national Follow-Through program. Washington, D. C.: Office of Education, U. S. Department of Health, Education, and Welfare, 1971. (b)

Stearns, M. S. Report on preschool programs: The effects of preschool programs on disadvantaged children and their families. Washington, D. C.: Office of Child Development, U. S. Department of Health, Education, and Welfare, 1971.

Westinghouse Learning Corporation. The impact of Head start experience: An evaluation of the effects of Head Start on children's cognitive and affective development. Vol. 1. Text and Appendices A-E. Ohio University Report to Office of Educational Opportunity, Clearinghouse for Federal, Scientific, and Technical Information, Washington, D. C., 1969.

Zigler, E., Abelson, J. D., Seitz, V. Motivational factors in the performance of economically disadvantaged children on the Peabody Picture Vocabulary Test. Child Development, 1973, 44, 294-303.

Acknowledgments

This research has been supported by Research Grant MH-05008 from the National Institutes of Mental Health, U. S. Public Health Service; by Research Grant 90-C-912 from the Office of Child Development, U. S. Department of Health, Education, and Welfare; Grant Number 1876-08830 from the Education Commission of the States; the Gunnar Dybwad Award of the National Association for Retarded Citizens; and the Follow Through Project of the U. S. Office of Education. The authors wish to express their appreciation to the staffs of the Hamden New Haven Cooperative Education Center and the Hamden and New Haven Public Schools for their cooperation and assistance.

Many persons have generously contributed towards making the post-intervention follow-up research possible. In particular, we wish to thank Robert S. Avery, Director of the Hamden-New Haven Cooperative Education Center; Audrey P. Tiani, Project Coordinator, Project Follow Through; Samuel Nash, Director of Special Projects and Planning for New Haven Public Schools; Jessie G. Bradley, Director of Elementary Schools, New Haven Public Schools; Gerald N. Tirozzi, Director of Middle Schools, New Haven Public Schools; Edward Mas, Superintendent, Hamden Public Schools; James J. Sullivan, Director of Parochial Schools, New Haven; Grayce Dowdy, Director of Head Start Programs, New Haven; Elizabeth Levine of the New Haven Prekindergarten Program; and the many principals, teachers, guidance counselors and others who assisted in the location of the children in these sam-

ples. We also thank Elizabeth Katz, who tested the Cohort 2 children in third and fourth grade; Sammy Carr and Cheryl LaRossa, who assisted extensively with data preparation; and Karen Anderson, who aided immeasurably with scoring and computer analyses; Bridget Thomsen, for clerical assistance with the manuscript, Jane Edmister, for assistance in preparation of graphics for this report; and Linda Hughes, who ably typed the manuscript.

Table 1

Differences Between Follow Through and Non Follow Through Children During the Kindergarten Year

		Low Income				Middle Income	
		Boys		Girls		Boys	Girls
		black & white	black only	black & white	black only		
Age in Months	FT	62.0	61.7	62.6	62.4	62.4	62.3
	NFT	62.1	62.1	63.7	64.1	64.5	64.2
Percent w/ HS or Nursery	FT	76***	83***	72***	80***	67	50
	NFT	35	35	30	39	54	41
Percent from single parent home	FT	30	29	32	27	11	8
	NFT	20	13	19	18	6	12
No. of Siblings	FT	2.8	2.9	3.0	3.2	2.0	2.5
	NFT	2.6	2.8	2.7	2.8	1.7	1.9
Fathers educ. in Years	FT	10.6	10.7	10.1	10.3	15.4	14.7
	NFT	10.0	9.6	9.5	9.8	14.8	15.1
Mothers educ. in Years	FT	11.5**	11.5*	10.8	11.0	13.8	13.6
	NFT	10.3	10.1	10.3	10.4	13.4	13.8
Star IQ KG Start	FT	90.7	89.6	93.4	90.6	105.1	108.6
	NFT	94.4	92.3	89.7	87.2	108.9	109.3
Star IQ KG End	FT	96.6	95.5	99.4	96.7	111.6	120.2
	NFT	98.9	94.4	94.0	92.0	111.1	115.9
PPVT IQ KG Start	FT	85.3	82.7	82.2	78.1	105.2	96.7
	NFT	82.0	77.5	73.0	69.8	109.2	98.7
PPVT IQ KG End	FT	95.4	94.3	90.2	88.6	112.2**	105.9
	NFT	96.6	93.0	90.1	88.5	124.0	104.0

*** The FT-NFT difference is significant, $p < .001$

** $p < .01$

* $p < .05$

Table 2

Comparison of Children Who Completed Follow Through With Children Who Left Follow Through (Cohort 1 Only)

	Boys				Girls			
	Black & White		Black Only		Black & White		Black Only	
	Yes[1]	No	Yes	No	Yes	No	Yes	No
Age in Months	---		---		---		---	
Percent with HS or Nursery	95 (N=21)	50** (N=16)	---		---		---	
Percent from Single Parent Home	---		---		---		---	
No. of Siblings	---		---		---		---	
Fathers Educ. in Years	---		---		---		---	
Mothers Educ. in Years	---		11.9 (N=17)	10.8* (N= 6)	---		---	
Star IQ, KG Start	---		---		---		---	
Star IQ, KG End	---		---		---		---	
PPVT IQ, KG Start	---		---		---		---	
PPVT IQ, KG End	---		99.8 (N=18)	86.6* (N= 5)	---		---	
PPVT IQ, Grade 1 End	---		---		---		95.1 (N=14)	81.2* (N= 5)
MET Wd. Kn, Grade 1	---		---		---		---	
MET Wd. Dis, Grade 1	---		---		---		---	
MET Reading, Grade 1	---		---		---		---	
MET Arithmetic, Grade 1	---		---		---		---	
MET Wd. Kn, Grade 2	---		---		---		---	
MET Wd. Dis, Grade 2	---		---		---		---	
MET Reading, Grade 2	---		---		---		---	
MET Arithmetic, Grade 2	---		---		---		---	

** The difference between survivors and nonsurvivors is significant, $p < .01$

* $p < .05$

---

[1] "Yes" denotes children who completed the FT program; "No" denotes children who did not.

Table 3

Comparison of Non-Follow Through Children Who Remained in Longitudinal Sample With Those Who Were Lost (Cohort 1 Only)

	Boys				Girls			
	Black & White		Black Only		Black & White		Black Only	
	Yes[1]	No	Yes	No	Yes	No	Yes	No
Age in Months	---		---		---		---	
Percent with HS or Nursery	---		---		---		---	
Percent from Single Parent Home	---		---		---		---	
No. of Siblings	---		---		---		---	
Fathers Educ. in Years	---		---		---		---	
Mothers Educ. in Years	---		---		---		---	
Star IQ, KG Start	---		---		96.1 (N=14)	80.4* (N= 8)	94.5 (N=10)	76.0* (N= 7)
Star IQ, KG End	---		---		---		---	
PPVT IQ, KG Start	---		---		---		---	
PPVT IQ, KG End	---		---		---		---	
PPVT IQ, Grade 1 End	---		---		---		---	
MET Wd. Kn, Grade 1	---		---		---		---	
MET Wd. Dis, Grade 1	---		---		---		---	
MET Reading, Grade 1	1.5 (N= 8)	1.1* (N= 2)	1.5 (N= 5)	1.1* (N= 2)	---		---	
MET Arithmetic, Grade 1	---		---		---		---	
MET Wd. Kn, Grade 2	---		---		---		---	
MET Wd. Dis, Grade 2	---		---		---		---	
MET Reading, Grade 2	---		---		---		---	
MET Arithmetic, Grade 2	---		---		---		---	

* The difference between the survivors and nonsurvivors is significant, $p < .05$

[1] "Yes" denotes children who remained in the longitudinal sample; "No" denotes children who did not.

Table 4
Comparison of Children Who Completed Follow Through
With Children Who Left Follow Through (Cohort 2 Only)

	Boys				Girls			
	Black & White		Black Only		Black & White		Black Only	
	Yes[1]	No	Yes	No	Yes	No	Yes	No
Age in Months	---		60.8 (N=24)	63.5* (N=10)	---		---	
Percent with HS or Nursery	85 (N=26)	54* (N=13)	---		---		---	
Percent from Single Parent Home	---		---		---		---	
No. of Siblings	---		---		---		---	
Fathers Educ. in Years	---		---		---		---	
Mothers Educ. in Years	---		---		---		---	
Star IQ, KG Start	---		---		---		---	
Star IQ KG End	---		---		---		---	
PPVT IQ KG Start	77.7 (N=24)	91.5* (N=13)	---		---		---	
PPVT IQ KG End	---		---		---		---	
PPVT IQ, Grade 1 End	---		---		---		102.6 (N=10)	89.17* (N= 6)
MET Wd. Kn, Grade 1	1.6 (N=20)	1.9* (N=8)	---		---		---	
MET Wd. Dis, Grade 1	1.8 (N=20)	2.3* (N=8)	1.8 (N=18)	2.4* (N=6)	---		---	
MET Reading, Grade 1	---		---		---		---	
MET Arithmetic, Grade 1	1.6 (N=20)	2.0* (N=8)	---		---		---	
MET Wd. Kn, Grade 2	2.2 (N=22)	3.0* (N=4)	---		---		---	
MET Wd. Dis, Grade 2	---		---		---		---	
MET Reading, Grade 2	---		---		---		3.1 (N=10)	2.0* (N=3)
MET Arithmetic, Grade 2	2.5 (N=22)	3.8* (N=4)	2.4 (N=20)	3.7* (N=3)	---		---	

**The difference between survivors and nonsurvivors is significant, $p < .01$
*$p < .05$

[1]"Yes" denotes children who completed the FT program; "No" denotes children who did not.

Table 5

Mean Values on Demographic Variables and Early School Performance for the Cohort 1 Low Income Longitudinal Follow Through and Non-Follow Through Children

Measure	Boys		Girls	
	FT (N=21)	NFT (N=13)	FT (N=14)	NFT (N=14)
Age in Months	62.6	64.0	61.8	65.0
Percent with HS or Nursery	95[a]	23	86[a]	14
Percent from Single Parent Home	24	15	43	15
No. of Siblings	2.6	3.4	3.1	3.2
Fathers Education in Years	10.4	8.9	11.1	9.8
Mothers Education in Years	11.8[a,b]	9.7	11.2	10.3
Star IQ, KG Start	88.6	81.9	93.1	96.1
Star IQ, KG End	96.4	92.9	101.3	96.7
PPVT IQ, KG Start	90.8[a,b]	69.7	75.4	70.2
PPVT IQ, KG End	101.2[b]	96.5	88.4	84.8
PPVT IQ, Grade 1 End	103.5[b]	95.5[c]	94.2	85.1
MET Wd. Kn, Grade 1	1.7	1.8	1.6	1.6
MET Wd. Dis, Grade 1	1.8	1.8	1.8	1.7
MET Reading, Grade 1	1.6	1.7	1.7	1.5
MET Arithmetic, Grade 1	1.6	1.8	1.6	1.7
MET Wd. Kn, Grade 2	2.5	2.4	2.4	2.5
MET Wd. Dis, Grade 2	2.7	2.9	2.9	2.9
MET Reading, Grade 2	2.3	2.6	2.4	2.6
MET Arithmetic, Grade 2	2.7	3.0	2.5	2.9

[a] The difference between the Follow Through and the non-Follow Through children of the same sex is significant at $p < .05$ or better.

[b] The difference between the sexes in the FT samples is significant at $p < .05$ or better.

[c] The difference between the sexes in the NFT sample is significant at $p < .05$.

Table 6

Significance Levels of One-tailed t-test Comparisons

of FT and NFT Low Income Children in Cohort 1

	Grade When Tested		
	Third	Sixth	Eighth
	Boys		
	(N=21 FT) (N=12 NFT)	(N =17 FT) (N= 13 NFT)	(N=19 FT) (N=12 NFT)
Mathematics	.07	.12	.10
Reading Recognition	---[a]	---	---
Reading Comprehension	---	---	ns[b]
Spelling	---	---	---
General Information	.07	.007**	.11
Total PIAT	ns	ns	ns
PPVT IQ	.09	.03*	.09
	Girls		
	(N=14 FT) (N=14 NFT)	(N=14 FT) (N=14 NFT)	(N=14 FT) (N=12 NFT)
Mathematics	ns	ns	.06
Reading Recognition	ns	---	---
Reading Comprehension	ns	---	---
Spelling	ns	ns	ns
General Information	---	ns	ns
Total PIAT	ns	---	---
PPVT IQ	ns	ns	---

[a]Dashed line denotes that the NFT mean was greater than the FT mean, although not significantly so.

[b]ns denotes that the one-tailed t value is greater than .15.

Table 7

Significance Levels of One-tailed t-test Comparisons

of FT and NFT Low Income Children in Cohort 2

	Grade When Tested		
	Third	Fifth	Seventh
	Boys		
	(N=21 FT) (N=18 NFT)	(N=20 FT) (N=17 NFT)	(N=20 FT) (N=17 NFT)
Mathematics	---[a]	---	---
Reading Recognition	---	---	ns[b]
Reading Comprehension	ns	ns	---
Spelling	---	---	---
General Information	---	---	---
Total PIAT	---	---	---
PPVT IQ	---	ns	---
	Girls		
	(N=10 FT) (N=19 NFT)	(N=10 FT) (N=18 NFT)	(N=10 FT) (N=17 NFT)
Mathematics	.035*	.04*	.06
Reading Recognition	---	ns	ns
Reading Comprehension	---	ns	ns
Spelling	ns	.05*	ns
General Information	.01**	.002**	.06
Total PIAT	ns	.045*	.10
PPVT IQ	.01**	.025*	.09

[a] Dashed line denotes that the NFT mean was greater than the FT mean; the difference was significant in one case (mathematics in seventh grade, boys)

[b] ns denotes that the one-tailed t value is greater than .15.

Table 8

Mean Values on Demongraphic Variables and Early School Performance for the Follow Through Samples Studied Since Third Grade

Measure	Cohort 1		Cohort 2	
	Boys (N=21)	Girls (N=14)	Boys (N=21)	Girls (N=10)
Percent from Single Parent Home	24	43	24	10
No. of Siblings	2.6	3.1	3.0	2.4
Fathers Education in Years	10.4	11.1	11.0	10.3
Mothers Education in Years	11.8[b]	11.2	11.4	11.6
Star IQ, KG Start	88.6	93.1	93.3	91.1
Star IQ, KG End	96.4	101.3	94.4	105.0
PPVT IQ, KG Start	90.9[a,b]	75.4	78.9	88.9
PPVT IQ, KG End	101.2[a,b]	88.4	92.0	93.0
PPVT IQ, Grade 1 End	103.5[a,b]	94.2[c]	93.5[b]	103.6
MET Wd. Kn, Grade 1	1.7	1.7	1.5	1.9
MET Wd. Dis, Grade 1	1.8	1.8	1.8[b]	2.4
MET Reading, Grade 1	1.6[a]	1.7	1.3[b]	1.7
MET Arithmetic, Grade 1	1.6	1.7	1.5	2.0
MET Wd. Kn, Grade 2	2.5	2.4	2.1[b]	2.8
MET Wd. Dis, Grade 2	2.7	2.9	2.5	3.4
MET Reading, Grade 2	2.3	2.4[c]	2.4[b]	3.1
MET Arithmetic, Grade 2	2.7[a]	2.5	2.3	3.0

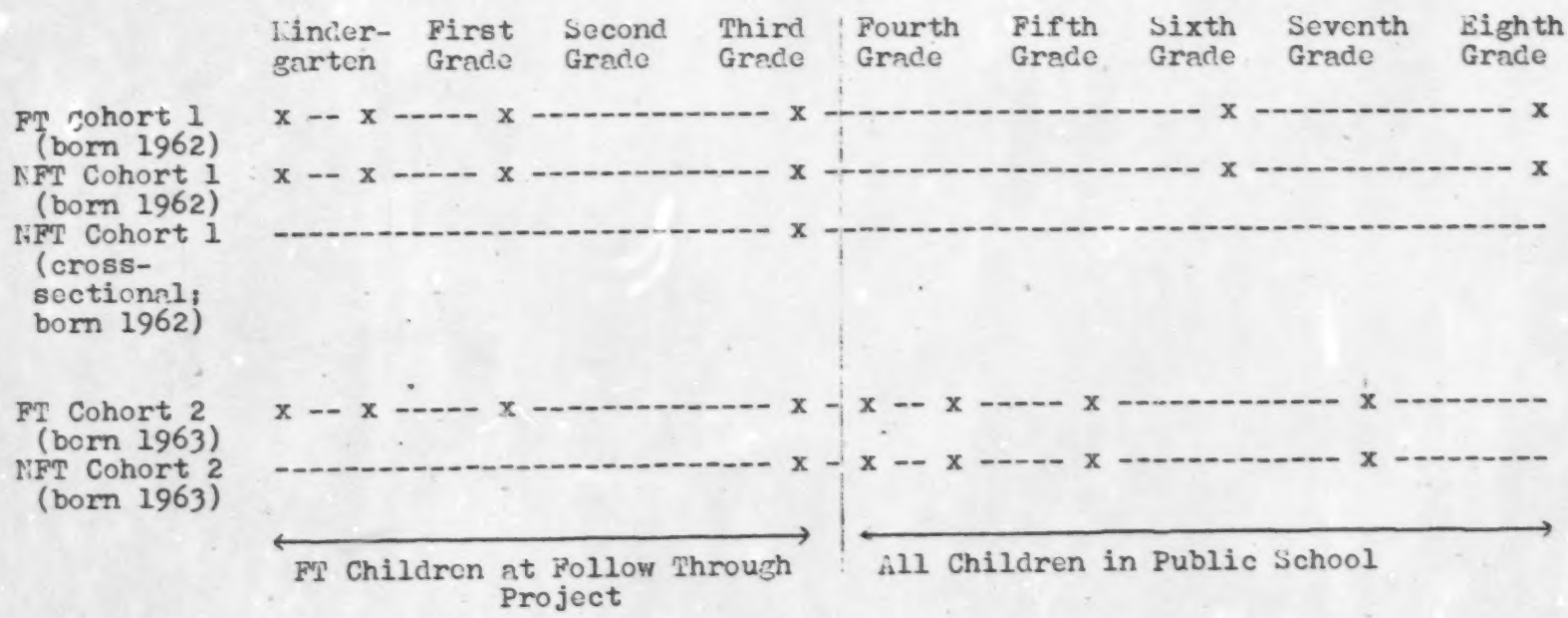
[a] Difference between the means of the boys in the two cohorts is significant, $p < .05$ or better.

[b] Difference between the sexes within the cohort is significant, $p < .05$ or better.

[c] Difference between the means of the girls in the two cohorts is significant, $p < .05$ or better.

Figure 1

A Summary of the Longitudinal Strategy for Evaluation of the Long-Term Effects of Intervention



x denotes time of testing

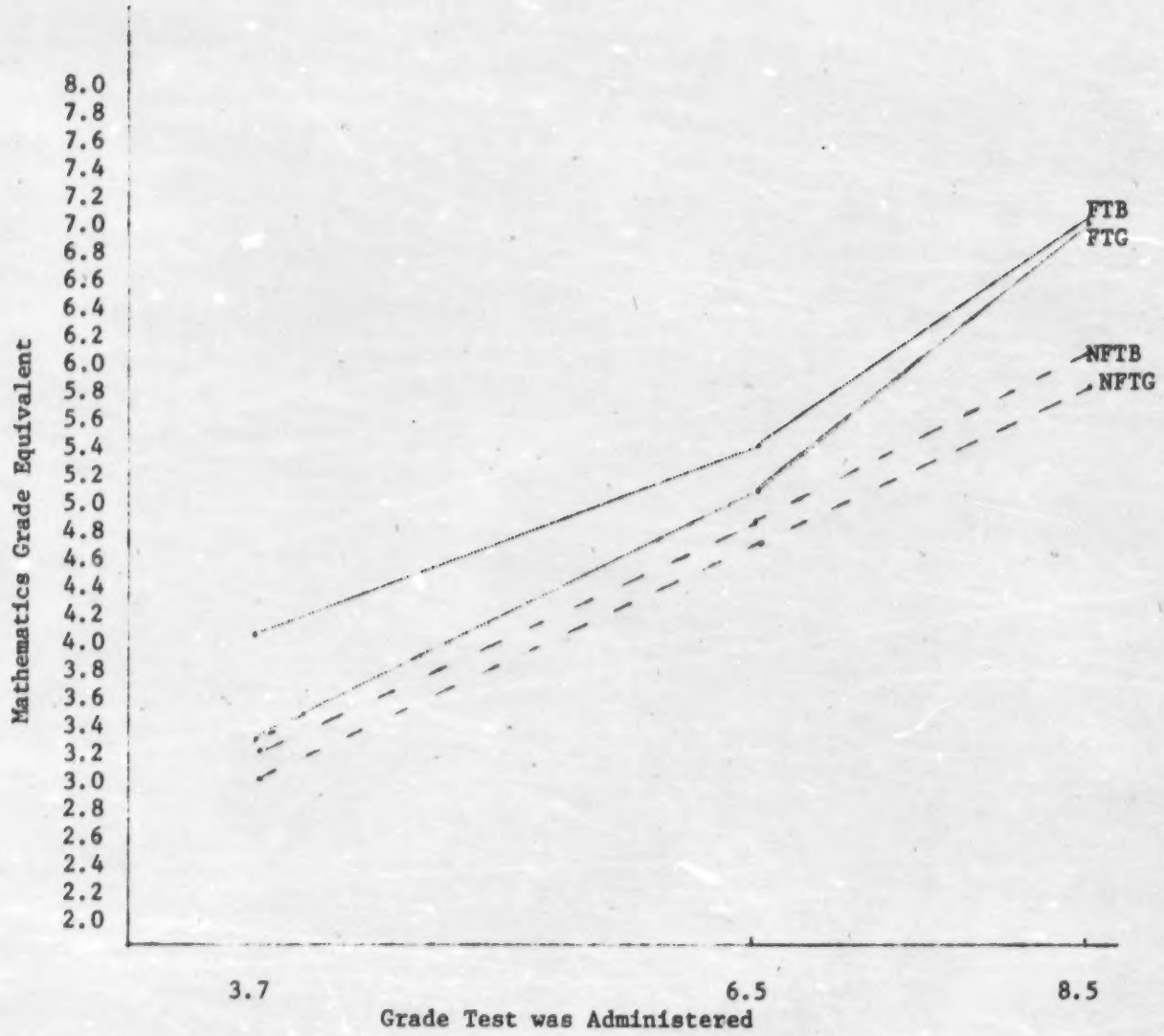


FIGURE 2

Cohort 1, Low Income Children
Mathematics Grade Equivalent Scores

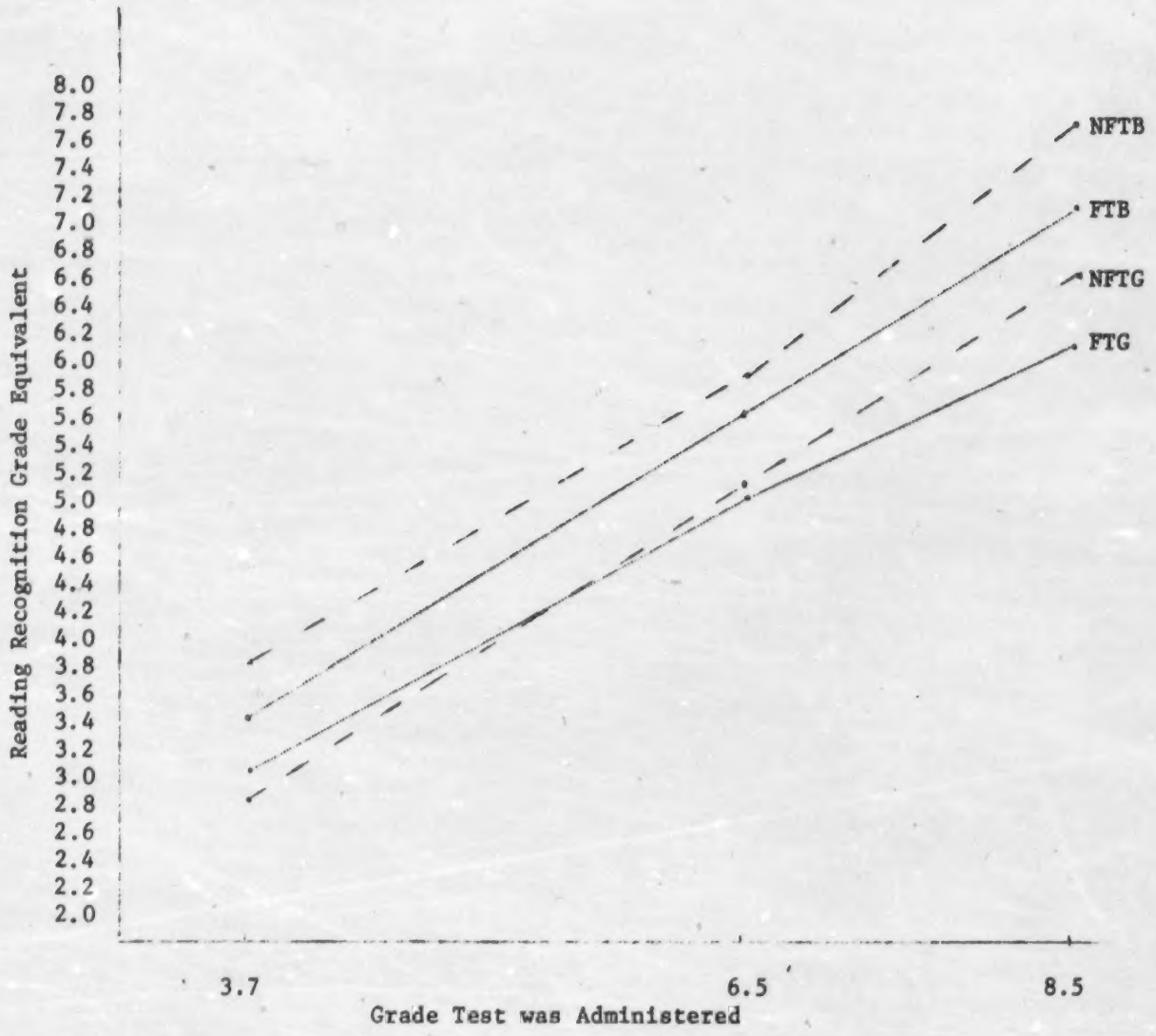


FIGURE 3

Cohort 1, Low Income Children
Reading Recognition Grade Equivalent Scores

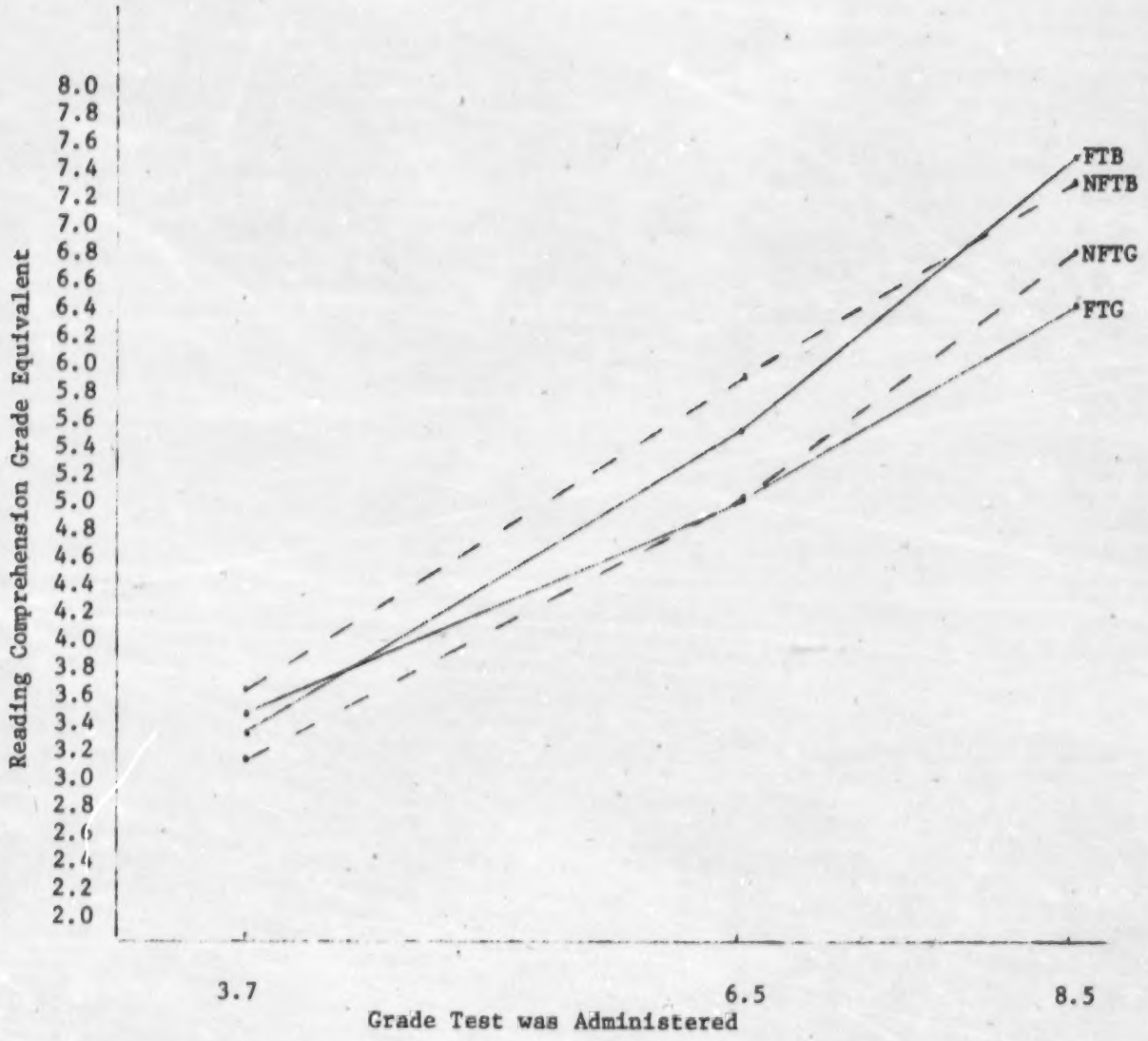


FIGURE 4

Cohort 1, Low Income Children
Reading Comprehension Grade Equivalent Scores

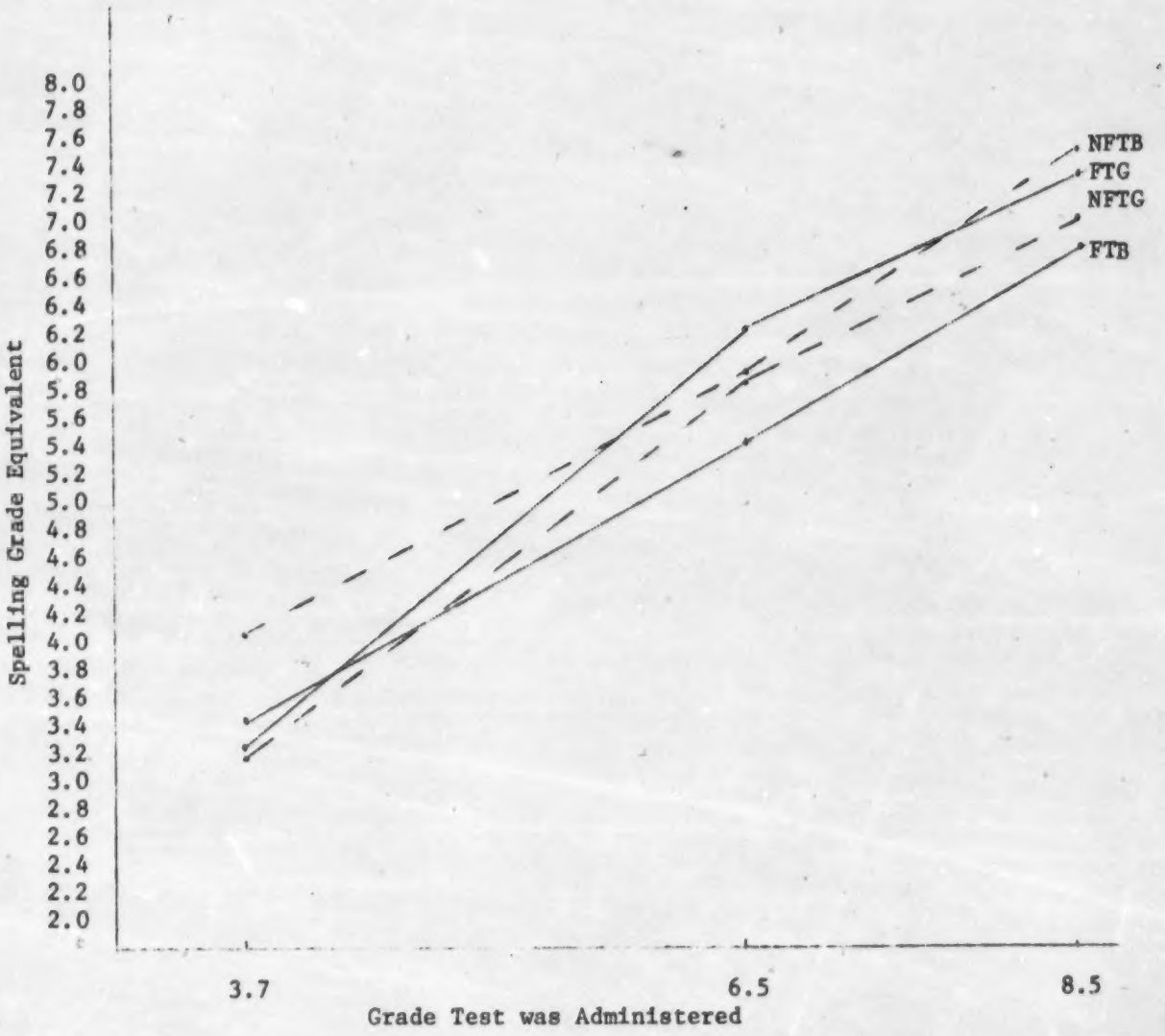


FIGURE 5

Cohort 1, Low Income Children
Spelling Grade Equivalent Scores

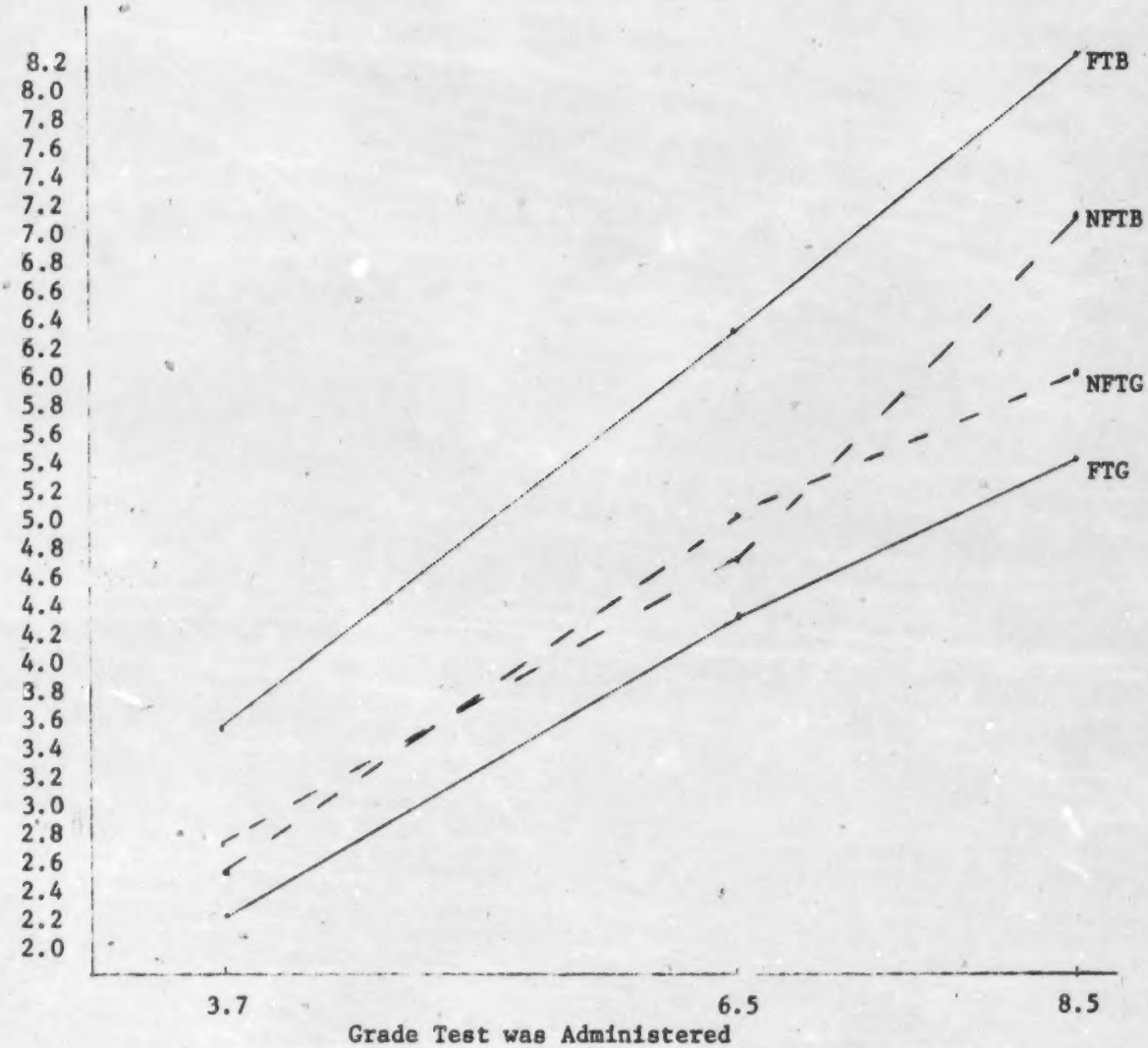


FIGURE 6

Cohort 1,Low Income Children
General Information Grade Equivalent Scores

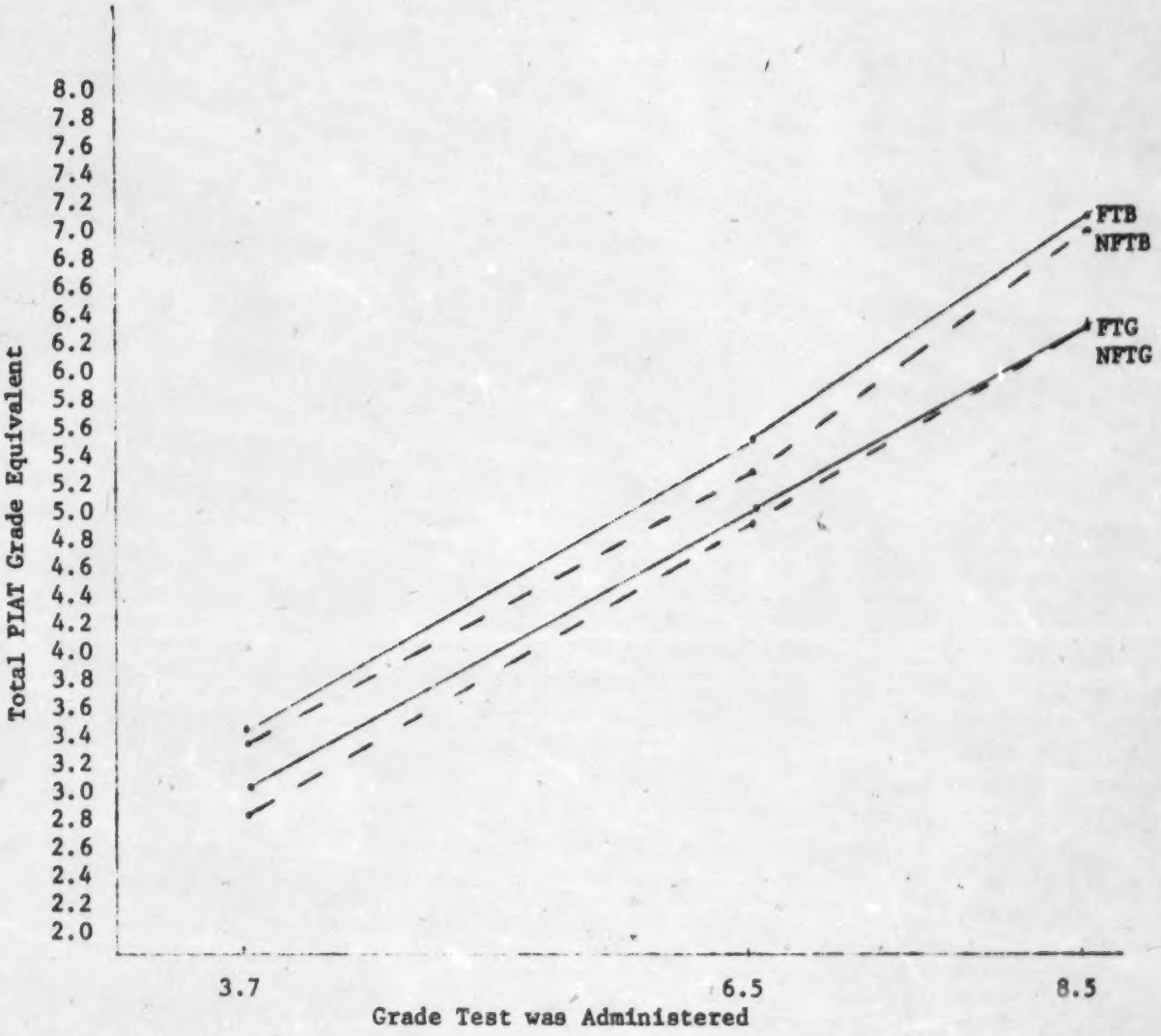


FIGURE 7

Cohort 1, Low Income Children
Total PIAT Grade Equivalent Scores

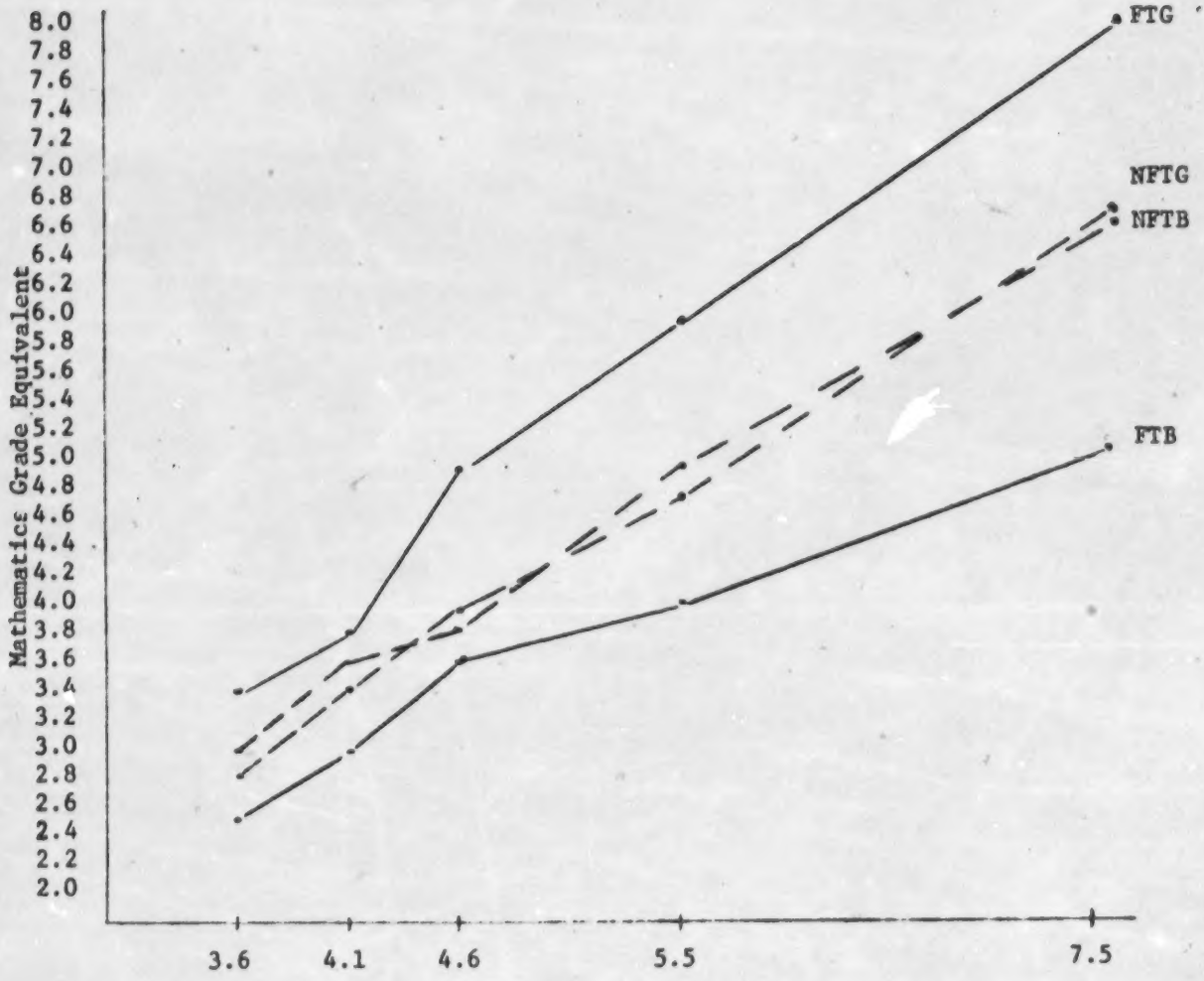


FIGURE 8

Cohort 2 Low Income Children
Mathematics Grade Equivalent Scores

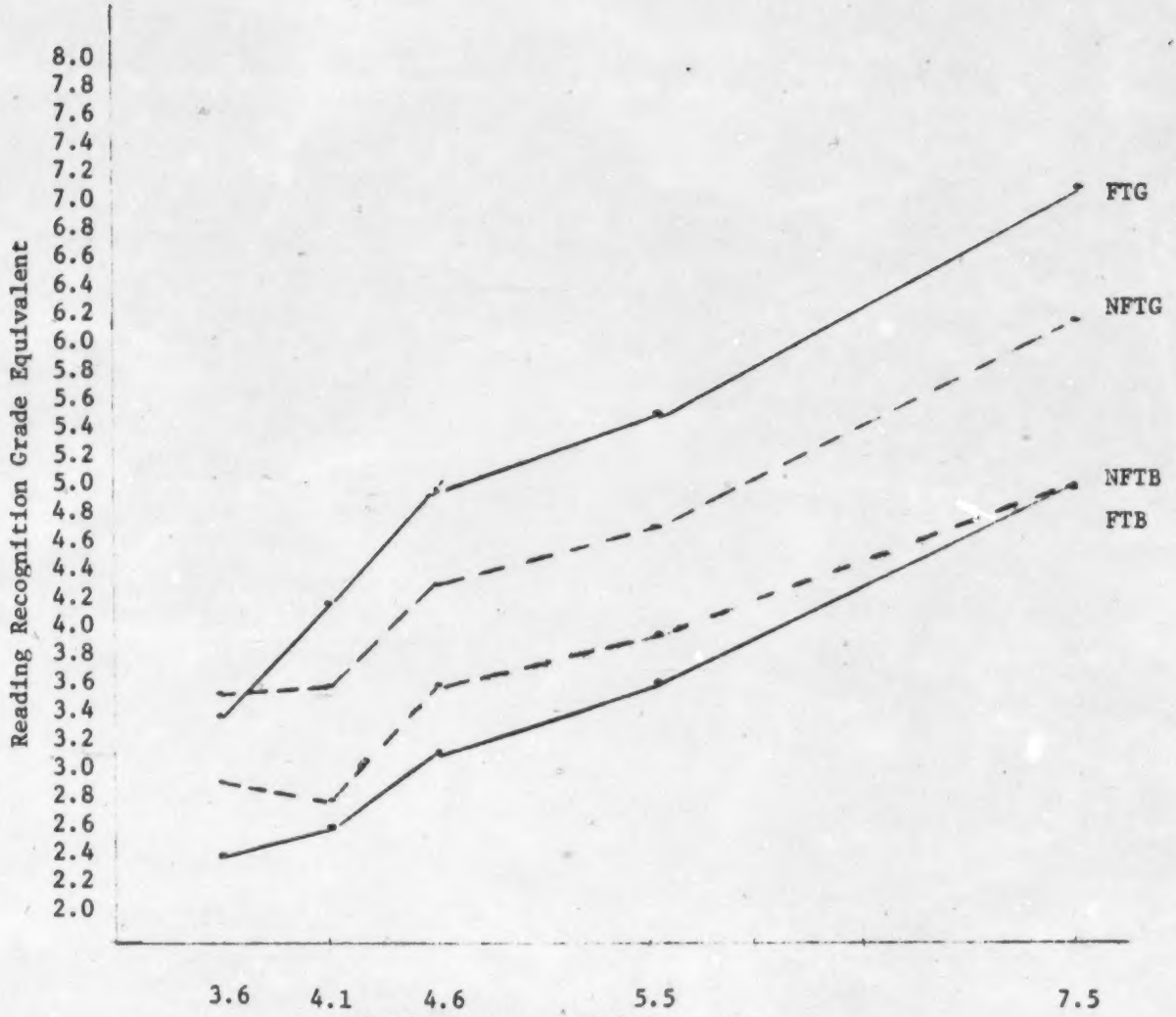


FIGURE 9

Cohort 2, Low Income Children
Reading Recognition Grade Equivalent Scores

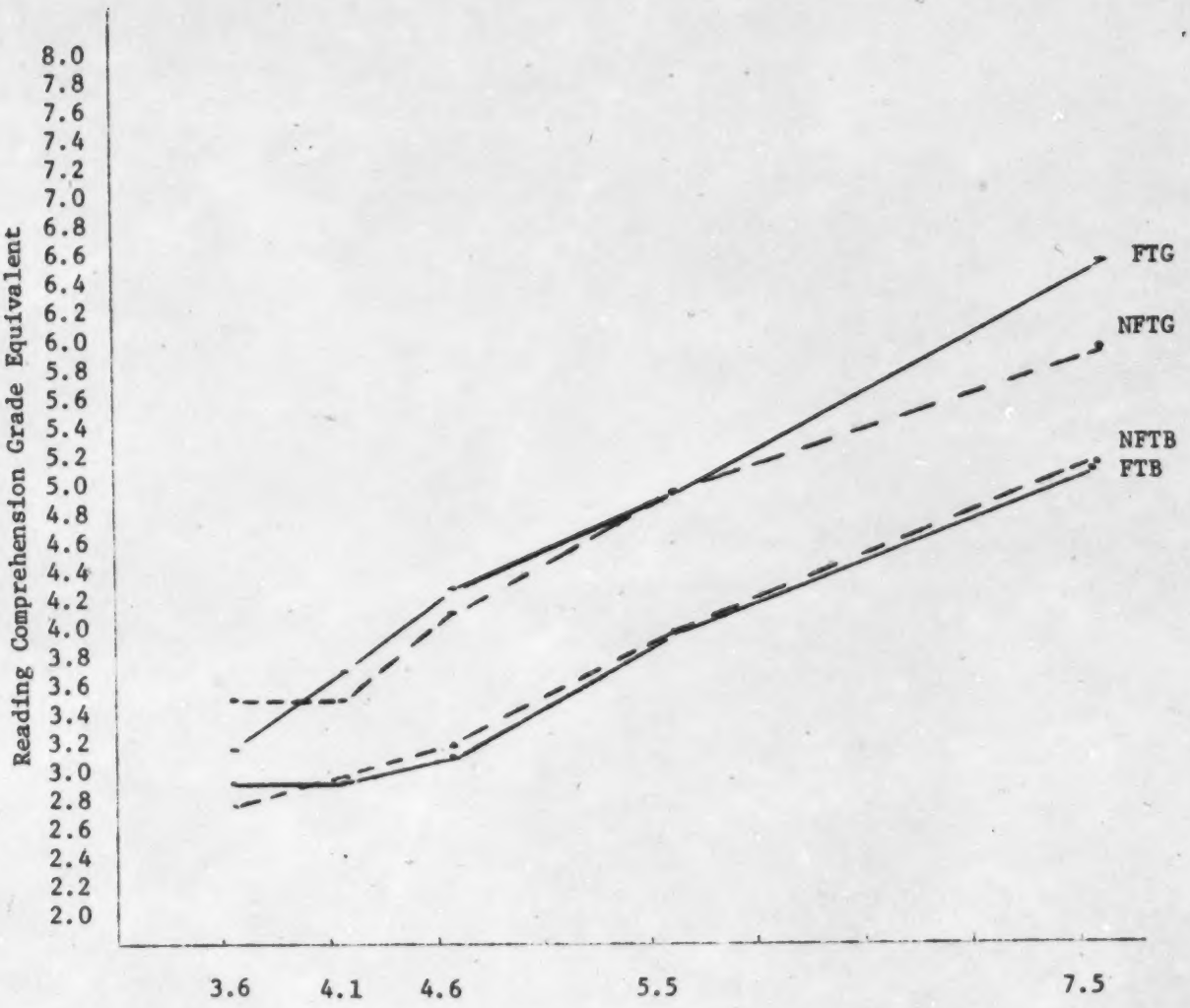


FIGURE 10

Cohort 2, Low Income Children
Reading Comprehension Grade Equivalent Scores

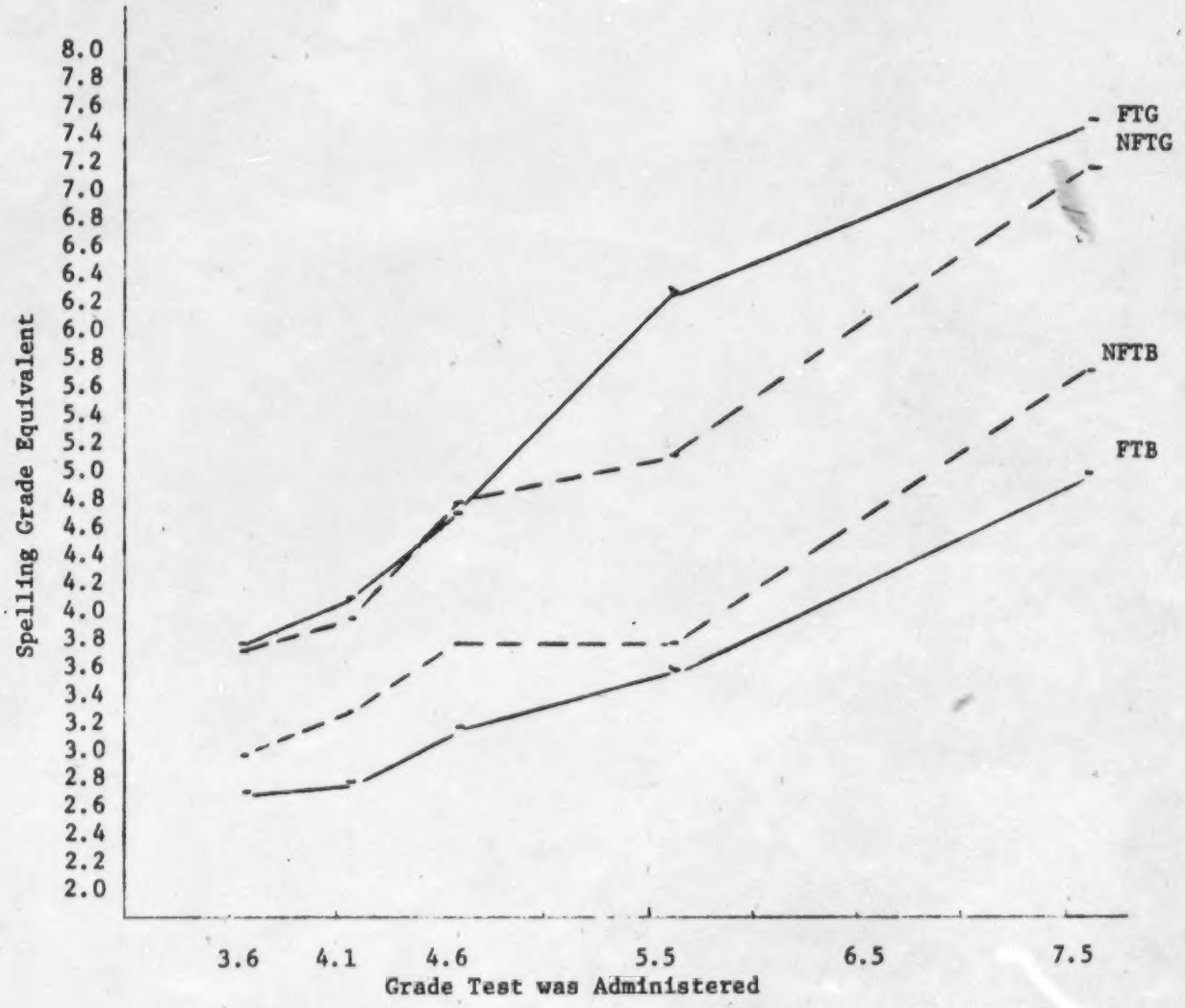


FIGURE 11

Cohort 2, Low Income Children
Spelling Grade Equivalent Scores

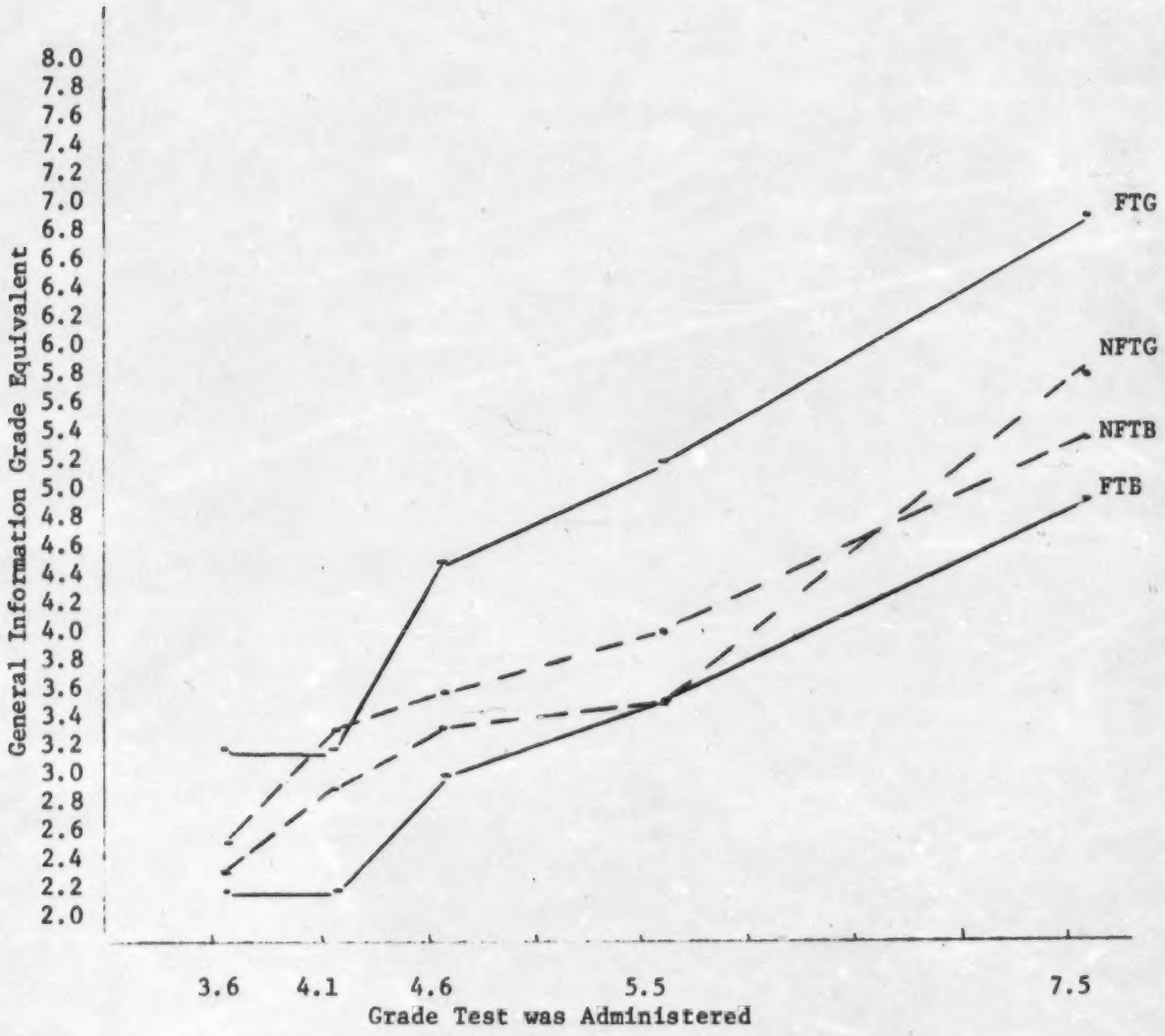


FIGURE 12

Cohort 2, Low Income Children
General Information Grade Equivalent Scores

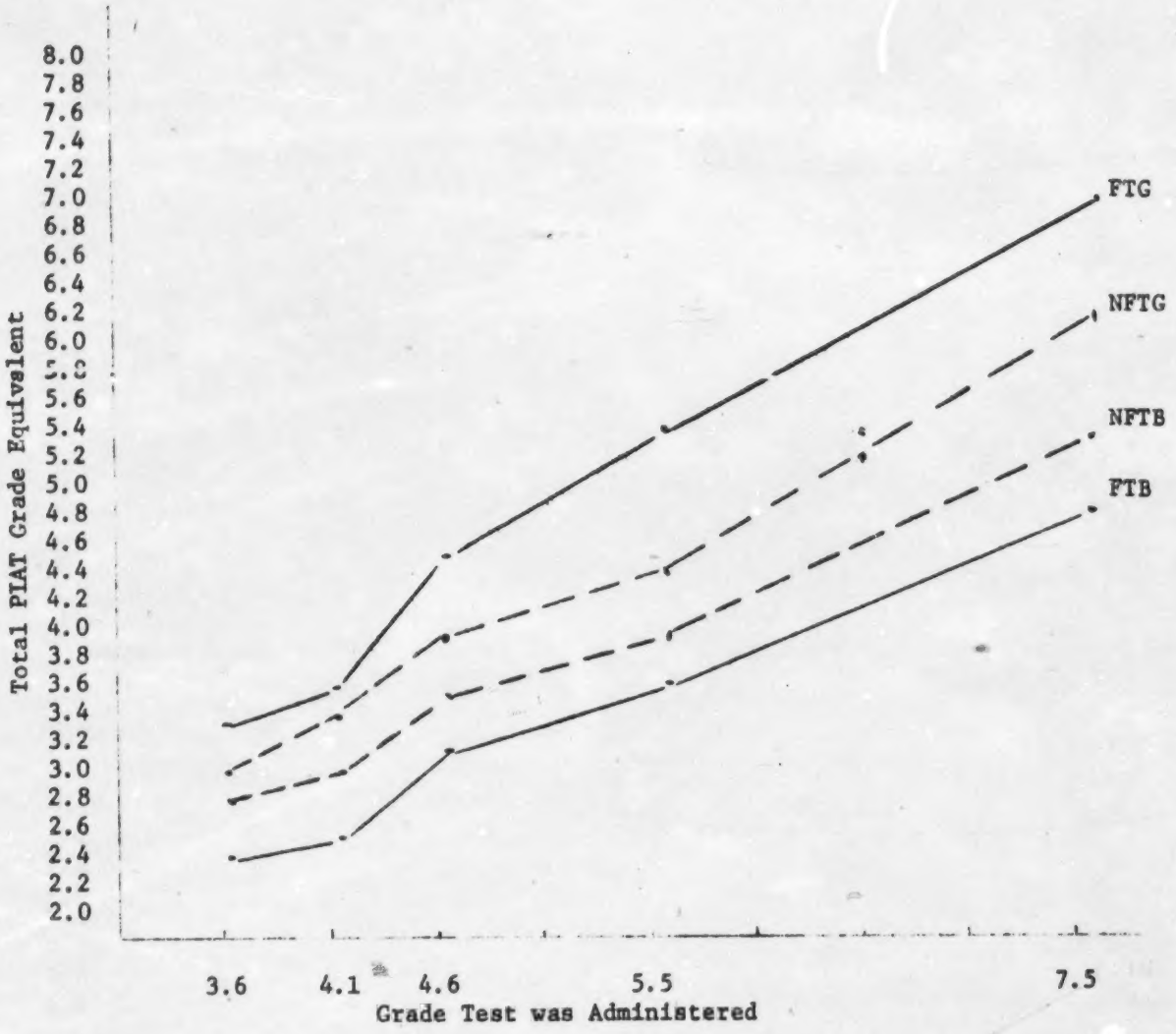


FIGURE 13

Cohort 2, Low Income Children
Total PIAT Grade Equivalent Scores